

# The resolvent average of monotone operators: dominant and recessive properties

Sedi Bartz\*, Heinz H. Bauschke†, Sarah M. Moffat‡ and Xianfu Wang§

May 11, 2015

## Abstract

Within convex analysis, a rich theory with various applications has been evolving since the proximal average of convex functions was first introduced over a decade ago. When one considers the subdifferential of the proximal average, a natural averaging operation of the subdifferentials of the averaged functions emerges. In the present paper we extend the reach of this averaging operation to the framework of monotone operator theory in Hilbert spaces, transforming it into the resolvent average. The theory of resolvent averages contains many desirable properties. In particular, we study a detailed list of properties of monotone operators and classify them as dominant or recessive with respect to the resolvent average. As a consequence, we recover a significant part of the theory of proximal averages. Furthermore, we shed new light on the proximal average and present novel results and desirable properties the proximal average possesses which have not been previously available.

**2010 Mathematics Subject Classification:** Primary 47H05, 52A41, 90C25; Secondary 15A09, 26A51, 26B25, 26E60, 47H09, 47A63.

**Keywords:** Convex function, Fenchel conjugate, Legendre function, monotone operator, para-monotone, positive semidefinite operator, proximal average, rectangular, resolvent, resolvent average, strong convexity, strong monotonicity, strong smoothness, subdifferential operator, uniform convexity, uniform smoothness.

---

\*Mathematics, Irving K. Barber School, University of British Columbia Okanagan, Kelowna, British Columbia V1V 1V7, Canada. E-mail: [sedi.bartz@ubc.ca](mailto:sedi.bartz@ubc.ca).

†Mathematics, Irving K. Barber School, University of British Columbia Okanagan, Kelowna, British Columbia V1V 1V7, Canada. E-mail: [heinz.bauschke@ubc.ca](mailto:heinz.bauschke@ubc.ca).

‡Mathematics, Irving K. Barber School, University of British Columbia Okanagan, Kelowna, British Columbia V1V 1V7, Canada. E-mail: [sarah.moffat@ubc.ca](mailto:sarah.moffat@ubc.ca).

§Mathematics, Irving K. Barber School, University of British Columbia Okanagan, Kelowna, British Columbia V1V 1V7, Canada. E-mail: [shawn.wang@ubc.ca](mailto:shawn.wang@ubc.ca).

# 1 Introduction

The proximal average of two convex functions was first considered in [9]. Since then, in a series of papers, the definition of the proximal average was refined and its useful properties were studied and employed in various applications revealing a rich theory with promising potential for further evolution and applications. One of the latest forms of the proximal average we refer to in the present paper is given in Definition 1.2 below. Some other cornerstones in the study of the proximal average include: [6] where many useful properties and examples were presented, [8] where it was demonstrated that the proximal average defines a homotopy on the class of convex functions (unlike other, classical averages), and, also, a significant application [13] where the proximal average was employed in order to explicitly construct *autoconjugate* representations of monotone operators, also known as *self-dual* Lagrangians, the importance of which in variational analysis is demonstrated in detail in the monograph [23]. A recent application of the proximal average in the theory of *machine learning* is [35]. When subdifferentiating the proximal average, we obtain an averaging operation of the subdifferentials of the underlying functions (see equation (4) below).

Monotone operators are fundamentally important in analysis and optimization [1], [5], [15], [16], [18], [33], [34]. In the present paper, we analyze the resolvent average (see Definition 1.4 below), which significantly extends the above averaging operation of subdifferentials to the general framework of monotone operator theory. (See also [10], [36], and [26] for some earlier works on the resolvent average.) We present powerful general properties the resolvent average possesses and then focus on the study of more specific inheritance properties of the resolvent average. Namely, we go through a detailed list of attractive properties of monotone operators and classify them as *dominant* or *recessive* with respect to the resolvent average by employing the following notions:

**Definition 1.1 (inheritance of properties)** *Let  $C$  be a set and let  $I$  be an index set. Suppose that  $\mathcal{AV}\mathcal{E} : C^I \rightarrow C$ . Then a property  $(p)$  is said to be*

- (i) **dominant** with respect to  $\mathcal{AV}\mathcal{E}$  if for each  $(c_i) \in C^I$ , the existence of  $i_0 \in I$  such that  $c_{i_0}$  has property  $(p)$  implies that  $\mathcal{AV}\mathcal{E}((c_i))$  has property  $(p)$ ;
- (ii) **recessive** with respect to  $\mathcal{AV}\mathcal{E}$  if  $(p)$  is not dominant and for each  $(c_i) \in C^I$ , for each  $i \in I$ ,  $c_i$  having property  $(p)$  implies that  $\mathcal{AV}\mathcal{E}((c_i))$  has property  $(p)$ .

We also provide several examples of the resolvent average (mainly in order to prove the recessive nature of several properties) of mappings which are monotone, however, which are not subdifferential operators. As a consequence, the resolvent average is now seen to be a natural and effective tool for averaging monotone operators which avoids many of the domain and range obstacles standing in front of classical averages such as the arithmetic average. The resolvent average is also seen to be an effective averaging technique when one wishes the average to possess specific properties, especially when the desired properties are dominant. When we restrict our attention to monotone linear relations, our current study extends the one in [12] where the resolvent average was considered as an average of positive semidefinite and definite matrices. When we restrict our attention to

subdifferential operators, we recover a large part of the theory of the proximal average [6]. Moreover, we present several novel results regarding the inheritance of desired properties of the proximal average which have not been previously available. In summary, *the resolvent average provides a novel technique for generating new maximally monotone operators with desirable properties.*

The remaining of the paper is organized as follows: In the remainder of Section 1 we present the basic definitions, notations and the relations between them which we will employ throughout the paper. We also collect all preliminary facts necessary for our presentation. In Section 2 we present basic properties of the resolvent average. In Section 3 we study dominant properties while Section 4 deals with recessive properties. Finally, in Section 5 we consider combinations of properties, properties which are neither dominant nor recessive, and other observations and remarks.

Before we start our analysis, let us recall the following concepts and standard notation from monotone operator theory and convex analysis: Throughout this paper,  $\mathcal{H}$  is a real Hilbert space with inner product  $\langle \cdot, \cdot \rangle$ , induced norm  $\|\cdot\|$ , identity mapping  $\text{Id}$  and we set  $q = \frac{1}{2}\|\cdot\|^2$ . We denote the interior of a subset  $C$  of  $\mathcal{H}$  by  $\text{int } C$ . Let  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  be a set-valued mapping. We say that  $A$  is *proper* when the *domain* of  $A$ , the set  $\text{dom } A = \{x \in \mathcal{H} \mid Ax \neq \emptyset\}$ , is nonempty. The *range* of  $A$  is the set  $\text{ran } A = A(\mathcal{H}) = \bigcup_{x \in \mathcal{H}} Ax$ , the *graph* of  $A$  is the set  $\text{gra } A = \{(x, u) \in \mathcal{H} \times \mathcal{H} \mid u \in Ax\}$  and the inverse of  $A$  is the mapping  $A^{-1}$  satisfying  $x \in A^{-1}u \Leftrightarrow u \in Ax$ .  $A$  is said to be *monotone* if

$$(\forall (x, u) \in \text{gra } A)(\forall (y, v) \in \text{gra } A) \quad \langle x - y, u - v \rangle \geq 0.$$

$A$  is said to be *maximally monotone* if there exists no monotone operator  $B$  such that  $\text{gra } A$  is a proper subset of  $\text{gra } B$ . The *resolvent* of  $A$  is the mapping  $J_A = (A + \text{Id})^{-1}$ . We say that  $A$  is a *linear relation* if  $\text{gra } A$  is a linear subspace of  $\mathcal{H} \times \mathcal{H}$ .  $A$  is said to be a *maximally monotone linear relation* if  $A$  is both maximally monotone and a linear relation. The mapping  $T : \mathcal{H} \rightarrow \mathcal{H}$  is said to be *firmly nonexpansive* if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|Tx - Ty\|^2 + \|(\text{Id} - T)x - (\text{Id} - T)y\|^2 \leq \|x - y\|^2.$$

Obviously, if  $T$  is firmly nonexpansive, then it is *nonexpansive*, that is, Lipschitz continuous with constant 1, where a Lipschitz continuous mapping with constant  $L$  is a mapping  $T : \mathcal{H} \rightarrow \mathcal{H}$  such that

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|Tx - Ty\| \leq L\|x - y\|.$$

The mapping  $T$  is said to be a *Banach contraction* if it is Lipschitz continuous with constant  $L < 1$ . The point  $x \in \mathcal{H}$  is said to be a *fixed point* of the mapping  $T : \mathcal{H} \rightarrow \mathcal{H}$  if  $Tx = x$ . The set of all fixed points of  $T$  is denoted by  $\text{Fix } T$ . The function  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is said to be *proper* if  $\text{dom } f = \{x \in \mathcal{H} \mid f(x) < \infty\} \neq \emptyset$ . The *Fenchel conjugate* of the function  $f$  is the function  $f^*$  which is defined by  $f^*(u) = \sup_{x \in \mathcal{H}} (\langle u, x \rangle - f(x))$ . The subdifferential of a proper function  $f$  is the mapping  $\partial f : \mathcal{H} \rightrightarrows \mathcal{H}$  which is defined by

$$\partial f(x) = \{u \in \mathcal{H} \mid f(x) + \langle u, y - x \rangle \leq f(y), \quad \forall y \in \mathcal{H}\}.$$

The *indicator function* of a subset  $C$  of  $\mathcal{H}$  is the function  $\iota_C : \mathcal{H} \rightarrow ]-\infty, +\infty]$  which vanishes on  $C$  and equals  $\infty$  on  $\mathcal{H} \setminus C$ . The *normal cone operator* of the set  $C$  is the mapping  $N_C = \partial \iota_C$ . We will denote the *nearest point projection* on the set  $C$  by  $P_C$ .

We now recall the definition of the proximal average and present the definition of the resolvent average. To this end we will make use of the following additional notations: Throughout the paper we assume that  $\mu \in ]0, \infty[$ ,  $n \in \{1, 2, \dots\}$  and  $I = \{1, \dots, n\}$ . For every  $i \in I$ , let  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  be a mapping, let  $f_i : \mathcal{H} \rightarrow ]-\infty, +\infty]$  be a function and let  $\lambda_i > 0$ ,  $\sum_{i \in I} \lambda_i = 1$ . We set:

$$\mathbf{A} = (A_1, \dots, A_n), \quad \mathbf{A}^{-1} = (A_1^{-1}, \dots, A_n^{-1}), \quad \mathbf{f} = (f_1, \dots, f_n), \quad \mathbf{f}^* = (f_1^*, \dots, f_n^*),$$

and  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)$ .

**Definition 1.2 (proximal average)** *The  $\boldsymbol{\lambda}$ -weighted proximal average of  $\mathbf{f}$  with parameter  $\mu$  is the function  $p_\mu(\mathbf{f}, \boldsymbol{\lambda}) : \mathcal{H} \rightarrow [-\infty, +\infty]$  defined by*

$$p_\mu(\mathbf{f}, \boldsymbol{\lambda})(x) = \frac{1}{\mu} \left( -\frac{1}{2} \|x\|^2 + \inf_{\sum_{i \in I} x_i = x} \sum_{i \in I} \lambda_i \left( \mu f_i(x_i / \lambda_i) + \frac{1}{2} \|x_i / \lambda_i\|^2 \right) \right), \quad x \in \mathcal{H}. \quad (1)$$

We will simply write  $p(\mathbf{f}, \boldsymbol{\lambda})$  when  $\mu = 1$ ,  $p_\mu(\mathbf{f})$  when all of the  $\lambda_i$ 's coincide and, finally,  $p(\mathbf{f})$  when  $\mu = 1$  and all of the  $\lambda_i$ 's coincide.

**Fact 1.3** [6, Proposition 4.3 and Theorem 6.7] *Suppose that for each  $i \in I$ ,  $f_i : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is proper, lower semicontinuous and convex. Then  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is proper, lower semicontinuous, convex and for every  $x \in \mathcal{H}$ ,*

$$p_\mu(\mathbf{f}, \boldsymbol{\lambda})(x) = \inf_{\sum_{i \in I} \lambda_j y_j = x} \sum_{i \in I} \lambda_i f_i(y_i) + \frac{1}{\mu} \left( \left( \sum_{i \in I} \lambda_i q(y_i) \right) - q(x) \right), \quad x \in \mathcal{H}. \quad (2)$$

Furthermore,

$$J_{\mu \partial p_\mu(\mathbf{f}, \boldsymbol{\lambda})} = \sum_{i \in I} \lambda_i J_{\mu \partial f_i}. \quad (3)$$

We recall that  $J_{\partial f} = \text{Prox} f$  is Moreau's proximity operator (see [27]). Thus, we see that  $\partial p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  defines an averaging operation of the  $\partial f_i$ 's with the weights  $\lambda_i$  and parameter  $\mu$  in the following manner:

$$\partial p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = \left( \sum_{i \in I} \lambda_i (\partial f_i + \mu^{-1} \text{Id})^{-1} \right)^{-1} - \mu^{-1} \text{Id}. \quad (4)$$

We now extend the reach of the averaging operation defined in (4) and which is the subject matter of the present paper:

**Definition 1.4 (resolvent average)** *The  $\boldsymbol{\lambda}$ -weighted resolvent average of  $\mathbf{A}$  with parameter  $\mu$  is defined by*

$$\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) = \left( \sum_{i \in I} \lambda_i (A_i + \mu^{-1} \text{Id})^{-1} \right)^{-1} - \mu^{-1} \text{Id}. \quad (5)$$

We will simply write  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$  when  $\mu = 1$ ,  $\mathcal{R}_\mu(\mathbf{A})$  when all of the  $\lambda_i$ 's coincide and, finally,  $\mathcal{R}(\mathbf{A})$  when  $\mu = 1$  and all of the  $\lambda_i$ 's coincide.

The motivation for naming (5) the *resolvent average* stems from the equivalence between equation (3) and equation (4), that is, from the fact that equation (5) is equivalent to the equation

$$J_{\mu\mathcal{R}_{\mu}(\mathbf{A},\boldsymbol{\lambda})} = \sum_{i \in I} \lambda_i J_{\mu A_i}. \quad (6)$$

The parameter  $\mu$  has been useful in the study of the proximal average; in particular, when taking  $\mu \rightarrow \infty$  or  $\mu \rightarrow 0^+$  one obtains classical averages of functions (see [6]). We employ particular choices of the parameter  $\mu$  in applications of the proximal average in the present paper as well.

Suppose that for each  $i \in I$ ,  $f_i : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is proper, lower semicontinuous, convex and set  $A_i = \partial f_i$ . By combining Definition 1.4 together with equation (4) we see that

$$\partial p_{\mu}(\mathbf{f}, \boldsymbol{\lambda}) = \mathcal{R}_{\mu}(\partial \mathbf{f}, \boldsymbol{\lambda}). \quad (7)$$

We end this introductory section with the following collection of facts which we will employ in the remaining sections of the paper.

The next fact from [33] was originally presented in the setting of finite-dimensional spaces, however, along with its proof from [33], it holds in any Hilbert space.

**Fact 1.5 (resolvent identity)** [33, Lemma 12.14] *For any mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$ ,*

$$J_A = \text{Id} - J_{A^{-1}}. \quad (8)$$

**Fact 1.6** [33, Proposition 6.17] *Let  $C$  be a nonempty, closed and convex subset of  $\mathcal{H}$ . Then*

$$J_{N_C} = (\text{Id} + N_C)^{-1} = P_C.$$

**Fact 1.7** [5, Proposition 4.2] *Let  $T : \mathcal{H} \rightarrow \mathcal{H}$ . Then the following are equivalent:*

- (i)  *$T$  is firmly nonexpansive.*
- (ii)  *$\text{Id} - T$  is firmly nonexpansive.*
- (iii)  *$2T - \text{Id}$  is nonexpansive.*
- (iv)  *$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \ \|Tx - Ty\|^2 \leq \langle x - y, Tx - Ty \rangle.$*

**Corollary 1.8** *Suppose that for each  $i \in I$ ,  $T_i : \mathcal{H} \rightarrow \mathcal{H}$  is firmly nonexpansive. Then  $T = \sum_{i \in I} \lambda_i T_i$  is firmly nonexpansive.*

*Proof.* Employing Fact 1.7, for each  $i \in I$ , letting  $N_i = 2T_i - \text{Id}$ , we see that  $N_i$  is nonexpansive. Letting  $N = \sum_{i \in I} \lambda_i N_i$ , we see that  $N$  is nonexpansive. Consequently, the mapping  $T = \frac{1}{2}(N + \text{Id})$  is firmly nonexpansive. ■

**Fact 1.9** [5, Lemma 2.13] *For each  $i \in I$  let  $x_i$  and  $u_i$  be points in  $\mathcal{H}$  and  $\alpha_i \in \mathbb{R}$  be such that  $\sum_{i \in I} \alpha_i = 1$ . Then*

$$\left\langle \sum_{i \in I} \alpha_i x_i, \sum_{j \in I} \alpha_j u_j \right\rangle + \frac{1}{2} \sum_{(i,j) \in I \times I} \alpha_i \alpha_j \langle x_i - x_j, u_i - u_j \rangle = \sum_{i \in I} \alpha_i \langle x_i, u_i \rangle. \quad (9)$$

Consequently,

$$\left\| \sum_{i \in I} \alpha_i x_i \right\|^2 = \sum_{i \in I} \alpha_i \|x_i\|^2 - \frac{1}{2} \sum_{(i,j) \in I \times I} \alpha_i \alpha_j \|x_i - x_j\|^2. \quad (10)$$

**Corollary 1.10** *Suppose that for each  $i \in I$ ,  $N_i : \mathcal{H} \rightarrow \mathcal{H}$  is nonexpansive,  $T_i : \mathcal{H} \rightarrow \mathcal{H}$  is firmly nonexpansive and set  $N = \sum_{i \in I} \lambda_i N_i$  and  $T = \sum_{i \in I} \lambda_i T_i$ . Let  $x$  and  $y$  be points in  $\mathcal{H}$  such that  $\|Tx - Ty\|^2 = \langle x - y, Tx - Ty \rangle$ . Then  $T_i x - T_i y = Tx - Ty$  for every  $i \in I$ . As a consequence, the following assertions hold:*

- (i) *If there exists  $i_0 \in I$  such that  $T_{i_0}$  is injective, then  $T$  is injective.*
- (ii) *If  $x$  and  $y$  are points in  $\mathcal{H}$  such that  $\|Nx - Ny\| = \|x - y\|$ , then  $N_i x - N_i y = Nx - Ny$  for every  $i \in I$ .*
- (iii) [29, Lemma 1.4] *If  $\bigcap_{i \in I} \text{Fix } N_i \neq \emptyset$ , then  $\text{Fix } N = \bigcap_{i \in I} \text{Fix } N_i$ .*

*Proof.* By employing equality (10) and then the firm nonexpansiveness of each  $T_i$  we obtain

$$\begin{aligned} \langle x - y, Tx - Ty \rangle &= \|Tx - Ty\|^2 \\ &= \sum_{i \in I} \lambda_i \|T_i x - T_i y\|^2 - \frac{1}{2} \sum_{(i,j) \in I \times I} \lambda_i \lambda_j \|(T_i x - T_i y) - (T_j x - T_j y)\|^2 \\ &\leq \sum_{i \in I} \lambda_i \langle x - y, T_i x - T_i y \rangle - \frac{1}{2} \sum_{(i,j) \in I \times I} \lambda_i \lambda_j \|(T_i x - T_i y) - (T_j x - T_j y)\|^2 \\ &= \langle x - y, Tx - Ty \rangle - \frac{1}{2} \sum_{(i,j) \in I \times I} \lambda_i \lambda_j \|(T_i x - T_i y) - (T_j x - T_j y)\|^2. \end{aligned}$$

Thus, we see that  $\|(T_i x - T_i y) - (T_j x - T_j y)\|^2 = 0$  for every  $i$  and  $j$  in  $I$ . Hence,  $T_i x - T_i y = Tx - Ty$  for every  $i \in I$ . (i): Follows immediately. (ii): For each  $i \in I$  we suppose that  $T_i = \frac{1}{2}(N_i + \text{Id})$  so that  $T_i$  as well as the mapping  $T = \sum_{i \in I} \lambda_i T_i = \frac{1}{2}(N + \text{Id})$  are firmly nonexpansive. Then the equality  $\|Nx - Ny\|^2 = \|x - y\|^2$  implies that  $\|Tx - Ty\|^2 = \langle x - y, Tx - Ty \rangle$ . Consequently, we see that  $T_i x - T_i y = Tx - Ty$  for every  $i \in I$ , which, in turn, implies that  $N_i x - N_i y = Nx - Ny$  for every  $i \in I$ . (iii): The inclusion  $\text{Fix } N \supseteq \bigcap_{i \in I} \text{Fix } N_i$  is trivial. Now, suppose that  $x \in \text{Fix } N$  and let  $y \in \bigcap_{i \in I} \text{Fix } N_i \subseteq \text{Fix } N$ . Then  $Nx - Ny = x - y$ , in particular,  $\|Nx - Ny\| = \|x - y\|$ . Consequently, for each  $i \in I$ ,  $x - y = Nx - Ny = N_i x - N_i y = N_i x - y$ , which implies that  $x = N_i x$ , as asserted by (iii). ■

**Fact 1.11 (Minty's Theorem)** [25] (see also [5, Theorem 21.1]) *Let  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  be monotone. Then*

$$\text{gra } A = \{(J_A x, (\text{Id} - J_A)x) \mid x \in \text{ran}(\text{Id} + A)\}. \quad (11)$$

*Furthermore,  $A$  is maximally monotone if and only if  $\text{ran}(\text{Id} + A) = \mathcal{H}$ .*

**Fact 1.12** ([20] and [25].) *Let  $T : \mathcal{H} \rightarrow \mathcal{H}$  and let  $A : \mathcal{H} \rightrightarrows \mathcal{H}$ . Then the following assertions hold:*

- (i) *If  $T$  is firmly nonexpansive, then  $B = T^{-1} - \text{Id}$  is maximally monotone and  $J_B = T$ .*
- (ii) *If  $A$  is maximally monotone, then  $J_A$  has full domain, and is single-valued and firmly nonexpansive, and  $A = J_A^{-1} - \text{Id}$ .*

**Definition 1.13 (Fitzpatrick function)** [21] *With the mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  we associate the Fitzpatrick function  $F_A : \mathcal{H} \times \mathcal{H} \rightarrow ]-\infty, +\infty]$ , defined by*

$$F_A(x, v) = \sup_{(z, w) \in \text{gra } A} (\langle w, x \rangle + \langle v, z \rangle - \langle w, z \rangle), \quad (x, v) \in \mathcal{H} \times \mathcal{H}. \quad (12)$$

**Definition 1.14 (rectangular and paramonotone mappings)** *The monotone mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is said to be rectangular [34, Definition 31.5] (also known as  $3^*$  monotone) if for every  $x \in \text{dom } A$  and every  $v \in \text{ran } A$  we have*

$$\inf_{(z, w) \in \text{gra } A} \langle v - w, x - z \rangle > -\infty, \quad (13)$$

*equivalently, if*

$$\text{dom } A \times \text{ran } A \subseteq \text{dom } F_A. \quad (14)$$

*The mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is said to be paramonotone if whenever we have a pair of points  $(x, v)$  and  $(y, u)$  in  $\text{gra } A$  such that  $\langle x - y, v - u \rangle = 0$ , then  $(x, u)$  and  $(y, v)$  are also in  $\text{gra } A$ .*

**Fact 1.15** [4, Remark 4.11], [14, Corollary 4.11] *Let  $A \in \mathbb{R}^{N \times N}$  be monotone and set  $A_+ = \frac{1}{2}A + \frac{1}{2}A^\top$ . Then the following assertions are equivalent:*

- (i)  *$A$  is paramonotone;*
- (ii)  *$A$  is rectangular;*
- (iii)  *$\text{rank } A = \text{rank } A_+$ ;*
- (iv)  *$\text{ran } A = \text{ran } A_+$ .*

**Fact 1.16** [19, I.2.3 and I.4] *Let  $A, B$  be linear relations. Then  $A^{-1}$  and  $A + B$  are linear relations.*

## 2 Basic properties of the resolvent average

In this section we present several basic properties of the resolvent average. These will stand as the foundation of our entire discussion in the present paper and will be applied repeatedly.

## 2.1 The inverse of the resolvent average

We begin our discussion by recalling the following fact [6, Theorem 5.1]: suppose that for each  $i \in I$ ,  $f_i : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is proper, lower semicontinuous and convex. Then

$$(p_\mu(\mathbf{f}, \boldsymbol{\lambda}))^* = p_{\mu^{-1}}(\mathbf{f}^*, \boldsymbol{\lambda}). \quad (15)$$

Now for each  $i \in I$  we set  $A_i = \partial f_i$ . Then  $A_i^{-1} = \partial f_i^*$ . Thus, recalling equation (7), we see that equation (15) turns into

$$(\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}))^{-1} = \partial(p_\mu(\mathbf{f}, \boldsymbol{\lambda}))^* = \partial p_{\mu^{-1}}(\mathbf{f}^*, \boldsymbol{\lambda}) = \mathcal{R}_{\mu^{-1}}(\mathbf{A}^{-1}, \boldsymbol{\lambda}), \quad (16)$$

that is, we have an inversion formula for the resolvent average in the case where we average subdifferential operators. We now aim at extending this result into our, more general, framework of the present paper. To this end, we will need the following general property of resolvents:

**Proposition 2.1** *Let  $A : \mathcal{H} \rightrightarrows \mathcal{H}$ . Then*

$$(A + \mu^{-1} \text{Id})^{-1} = (\text{Id} - \mu(A^{-1} + \mu \text{Id})^{-1}) \circ (\mu \text{Id}) \quad (17)$$

$$= \mu(\text{Id} - (\mu^{-1}A^{-1} + \text{Id})^{-1}). \quad (18)$$

*Consequently, if  $B : \mathcal{H} \rightrightarrows \mathcal{H}$  is a mapping such that*

$$(A + \mu^{-1} \text{Id})^{-1} = (\text{Id} - \mu(B + \mu \text{Id})^{-1}) \circ (\mu \text{Id}), \quad (19)$$

*then  $B = A^{-1}$ .*

*Proof.* For any mapping  $F : \mathcal{H} \rightrightarrows \mathcal{H}$  we have  $(\mu F)^{-1} = F^{-1} \circ (\mu^{-1} \text{Id})$ . Employing this fact, the fact that  $A + \mu^{-1} \text{Id} = \mu^{-1}(\mu A + \text{Id})$  and the resolvent identity (8), we obtain the following chain of equalities:

$$\begin{aligned} (A + \mu^{-1} \text{Id})^{-1} &= J_{\mu A} \circ (\mu \text{Id}) = (\text{Id} - J_{(\mu A)^{-1}}) \circ (\mu \text{Id}) = (\text{Id} - J_{A^{-1} \circ (\mu^{-1} \text{Id})}) \circ (\mu \text{Id}) \\ &= \left( \text{Id} - (A^{-1} \circ (\mu^{-1} \text{Id}) + \text{Id})^{-1} \right) \circ (\mu \text{Id}) \\ &= \left( \text{Id} - (A^{-1} \circ (\mu^{-1} \text{Id}) + (\mu \text{Id}) \circ (\mu^{-1} \text{Id}))^{-1} \right) \circ (\mu \text{Id}) \\ &= \left( \text{Id} - ((A^{-1} + \mu \text{Id}) \circ (\mu^{-1} \text{Id}))^{-1} \right) \circ (\mu \text{Id}) = (\text{Id} - \mu(A^{-1} + \mu \text{Id})^{-1}) \circ (\mu \text{Id}) \\ &= \mu \text{Id} - \mu(\mu^{-1}A^{-1} + \text{Id})^{-1} \circ (\mu^{-1} \text{Id}) \circ (\mu \text{Id}) = \mu(\text{Id} - (\mu^{-1}A^{-1} + \text{Id})^{-1}). \end{aligned}$$

This completes the proof of (17) and (18). Now, suppose that  $B : \mathcal{H} \rightrightarrows \mathcal{H}$  is a mapping which satisfies equation (19). Then, by employing equation (17) to the right hand side of equation (19), we arrive at  $(A + \mu^{-1} \text{Id})^{-1} = (B^{-1} + \mu^{-1} \text{Id})^{-1}$ , which implies that  $A + \mu^{-1} \text{Id} = B^{-1} + \mu^{-1} \text{Id}$ . Since  $\mu^{-1} \text{Id}$  is single-valued, we conclude that  $B^{-1} = A$  and complete the proof.  $\blacksquare$

**Theorem 2.2 (the inversion formula)** *Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is a set-valued mapping. Then*

$$(\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}))^{-1} = \mathcal{R}_{\mu^{-1}}(\mathbf{A}^{-1}, \boldsymbol{\lambda}). \quad (20)$$



*Proof.* By employing the definition (see (5)) of  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$ ,  $\mathcal{R}_{\mu^{-1}}(\mathbf{A}^{-1}, \boldsymbol{\lambda})$  in (21), (23) and by also employing equation (17) in (22) below, we obtain the following chain of equalities:

$$\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) = \left( \sum_{i \in I} \lambda_i (A_i + \mu^{-1} \text{Id})^{-1} \right)^{-1} - \mu^{-1} \text{Id} \quad (21)$$

$$= \left( \sum_{i \in I} \lambda_i (\text{Id} - \mu(A_i^{-1} + \mu \text{Id})^{-1}) \circ (\mu \text{Id}) \right)^{-1} - \mu^{-1} \text{Id} \quad (22)$$

$$= \left( \mu \text{Id} - \mu \sum_{i \in I} \lambda_i (A_i^{-1} + \mu \text{Id})^{-1} \circ (\mu \text{Id}) \right)^{-1} - \mu^{-1} \text{Id}$$

$$= \left( \mu \text{Id} - \mu (\mathcal{R}_{\mu^{-1}}(\mathbf{A}^{-1}, \boldsymbol{\lambda}) + \mu \text{Id})^{-1} \circ (\mu \text{Id}) \right)^{-1} - \mu^{-1} \text{Id} \quad (23)$$

which, in turn, implies

$$(\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) + \mu^{-1} \text{Id})^{-1} = (\text{Id} - \mu[\mathcal{R}_{\mu^{-1}}(\mathbf{A}^{-1}, \boldsymbol{\lambda}) + \mu \text{Id}]^{-1}) \circ (\mu \text{Id}).$$

Finally, letting  $A = \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  and  $B = \mathcal{R}_{\mu^{-1}}(\mathbf{A}^{-1}, \boldsymbol{\lambda})$ , we may now apply characterization (19) in order to obtain  $A^{-1} = B$  and completes the proof.  $\blacksquare$

We see that, indeed, Theorem 2.2 extends the reach of formula (16). Consequently, since the subdifferential of a proper, lower semicontinuous and convex function determines its antiderivative uniquely up to an additive constant, we note that, in fact, Theorem 2.2 recovers formula (15) up to an additive constant.

## 2.2 Basic properties of $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$ and common solutions to monotone inclusions

**Proposition 2.3** *Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is maximally monotone.*

(i) *Let  $x$  and  $u$  be points in  $\mathcal{H}$ . Then*

$$\mathcal{R}_\mu((A_1 - u, \dots, A_n - u), \boldsymbol{\lambda}) = \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) - u \quad (24)$$

*and*

$$\mathcal{R}_\mu((A_1((\cdot) - x), \dots, A_n((\cdot) - x)), \boldsymbol{\lambda}) = \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})((\cdot) - x). \quad (25)$$

(ii) *Let  $0 < \alpha$ . Then*

$$\mathcal{R}_\mu(\alpha \mathbf{A}, \boldsymbol{\lambda}) = \alpha \mathcal{R}_{\alpha\mu}(\mathbf{A}, \boldsymbol{\lambda}) \quad \text{in particular,} \quad \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) = \mu^{-1} \mathcal{R}(\mu \mathbf{A}, \boldsymbol{\lambda}). \quad (26)$$

(iii) *Let  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  be maximally monotone and suppose that for each  $i \in I$ ,  $A_i = A$ . Then*

$$\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) = A. \quad (27)$$

*Proof.* (i): For a mapping  $B : \mathcal{H} \rightrightarrows \mathcal{H}$  we have  $(B - u)^{-1} = B^{-1}((\cdot) + u)$ . Thus,

$$J_{\mu \mathcal{R}_\mu((A_1 - u, \dots, A_n - u), \boldsymbol{\lambda})} = \sum_{i \in I} \lambda_i J_{\mu(A_i - u)} = \sum_{i \in I} \lambda_i J_{\mu A_i}((\cdot) + \mu u) = J_{\mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})}((\cdot) + \mu u).$$

Consequently, (24) follows. A similar argument also implies equation (25). (ii): Since  $J_{\mu \mathcal{R}_\mu(\alpha \mathbf{A}, \boldsymbol{\lambda})} = \sum_{i \in I} \lambda_i J_{\mu \alpha A_i} = J_{\mu \alpha \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})}$ , equation (26) follows. (iii): Since  $J_{\mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})} = \sum_{i \in I} \lambda_i J_{\mu A_i} = J_{\mu A}$ , we obtain  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) = A$ .  $\blacksquare$

**Remark 2.4** Suppose that for each  $i \in I$ ,  $f_i : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is proper, lower semicontinuous and convex and we set  $A_i = \partial f_i$ . In this particular case formula (26) can be obtained by subdifferentiating the following formula [6, Remark 4.2(iv)]:

$$p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = \mu^{-1} p(\mu \mathbf{f}, \boldsymbol{\lambda}). \quad (28)$$

A strong motivation for studying the resolvent average stems from the fact that it also captures common solutions to monotone inclusions:

**Theorem 2.5 (common solutions to monotone inclusions)** *Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is maximally monotone. Let  $x$  and  $u$  be points in  $\mathcal{H}$ . If  $\bigcap_{i \in I} A_i(x) \neq \emptyset$ , then*

$$\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})(x) = \bigcap_{i \in I} A_i(x). \quad (29)$$

*If  $\bigcap_{i \in I} A_i^{-1}(u) \neq \emptyset$ , then*

$$\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})^{-1}(u) = \bigcap_{i \in I} A_i^{-1}(u). \quad (30)$$

*Proof.* First we prove that if  $\bigcap_{i \in I} A_i^{-1}(0) \neq \emptyset$  then  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})^{-1}(0) = \bigcap_{i \in I} A_i^{-1}(0)$ . Our assumption that  $\bigcap_{i \in I} A_i^{-1}(0) \neq \emptyset$  means that  $\bigcap_{i \in I} \text{Fix } J_{A_i} \neq \emptyset$ . Since for each  $i \in I$ ,  $J_{A_i}$  is nonexpansive, Corollary 1.10(iii) guarantees that

$$\text{Fix } J_{\mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})} = \text{Fix } \sum_{i \in I} \lambda_i J_{\mu A_i} = \bigcap_{i \in I} \text{Fix } J_{\mu A_i}$$

which implies that  $(\mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}))^{-1}(0) = \bigcap_{i \in I} (\mu A_i^{-1})(0)$  and, consequently, that  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})^{-1}(0) = \bigcap_{i \in I} A_i^{-1}(0)$ . Now, let  $u \in \mathcal{H}$ . Given a mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$ , then  $A^{-1}(u) = (A - u)^{-1}(0)$ . Consequently, if  $\bigcap_{i \in I} A_i^{-1}(u) \neq \emptyset$ , then  $\bigcap_{i \in I} (A_i - u)^{-1}(0) \neq \emptyset$ . By employing equation (24) we obtain

$$\begin{aligned} \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})^{-1}(u) &= (\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) - u)^{-1}(0) = \mathcal{R}_\mu((A_1 - u, \dots, A_n - u), \boldsymbol{\lambda})^{-1}(0) \\ &= \bigcap_{i \in I} (A_i - u)^{-1}(0) = \bigcap_{i \in I} A_i^{-1}(u) \end{aligned}$$

which completes the proof of equation (30). Let  $x \in \mathcal{H}$ . If  $\bigcap_{i \in I} A_i(x) \neq \emptyset$ , then by employing Theorem 2.2 we now obtain

$$\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})(x) = (\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})^{-1})^{-1}(x) = (\mathcal{R}_{\mu^{-1}}(\mathbf{A}^{-1}, \boldsymbol{\lambda}))^{-1}(x) = \bigcap_{i \in I} (A_i^{-1})^{-1}(x) = \bigcap_{i \in I} A_i(x)$$

which completes the proof of equation (29). ■

**Example 2.6 (convex feasibility problem)** Suppose that for each  $i \in I$ ,  $C_i$  is a nonempty, closed and convex subset of  $\mathcal{H}$  and set  $A_i = N_{C_i}$ . If  $\bigcap_{i \in I} C_i \neq \emptyset$ , then

$$\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})^{-1}(0) = \bigcap_{i \in I} C_i.$$

### 2.3 Monotonicity, domain, range and the graph of the resolvent average

We continue our presentation of general properties of the resolvent average by focusing our attention on monotone operators.

**Theorem 2.7** *Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is a set-valued mapping. Then  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is maximally monotone if and only if for each  $i \in I$ ,  $A_i$  is maximally monotone. In this case*

$$\text{gra } \mu\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) \subseteq \sum_{i \in I} \lambda_i \text{gra } \mu A_i, \quad (31)$$

and, consequently,

$$\text{ran } \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) \subseteq \sum_{i \in I} \lambda_i \text{ran } A_i \quad \text{and} \quad \text{dom } \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) \subseteq \sum_{i \in I} \lambda_i \text{dom } A_i. \quad (32)$$

*Proof.* By employing equation (6) we see that

$$\text{dom } J_{\mu\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})} = \bigcap_{i \in I} \text{dom } J_{\mu A_i}. \quad (33)$$

As a consequence, we see that  $\text{ran}(\text{Id} + \mu\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})) = \mathcal{H}$  if and only if for each  $i \in I$ ,  $\text{ran}(\text{Id} + \mu A_i) = \mathcal{H}$ . Recalling Minty's Theorem (Fact 1.11), we see that  $\mu\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is maximally monotone if and only if for each  $i \in I$ ,  $\mu A_i$  is maximally monotone, that is,  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is maximally monotone if and only if for each  $i \in I$ ,  $A_i$  is maximally monotone. Finally, applying Minty's parametrization (11) we obtain

$$\begin{aligned} \text{gra } \mu\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) &= \left\{ (J_{\mu\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})}(x), x - J_{\mu\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})}x) \mid x \in \text{dom } J_{\mu\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})} \right\} \\ &= \left\{ \left( \sum_{i \in I} \lambda_i J_{\mu A_i} x, \sum_{i \in I} \lambda_i (\text{Id} - J_{\mu A_i}) x \right) \mid x \in \bigcap_{i \in I} \text{dom } J_{\mu A_i} \right\} \\ &= \left\{ \sum_{i \in I} \lambda_i (J_{\mu A_i} x, (\text{Id} - J_{\mu A_i}) x) \mid x \in \bigcap_{i \in I} \text{dom } J_{\mu A_i} \right\} \subseteq \sum_{i \in I} \lambda_i \text{gra } \mu A_i. \end{aligned}$$

which implies inclusion (31). ■

As an example, we now discuss the case where we average a monotone mapping with its inverse. It was observed in [6, Example 5.3] that when  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is proper, lower semicontinuous and convex, then

$$p(f, f^*) = q. \quad (34)$$

Employing equation (7) we see that

$$\mathcal{R}(\partial f, \partial f^*) = \partial p(f, f^*) = \partial q = \text{Id}.$$

We now extend the reach of this fact in order for it to hold in the framework of the resolvent average of monotone operators. To this end we first note the following fact:

**Proposition 2.8** *Let  $R : \mathcal{H} \rightrightarrows \mathcal{H}$  be a set-valued and monotone mapping such that  $R = R^{-1}$ . Then  $\text{gra } R \subseteq \text{gra Id}$ . Consequently, if  $R$  is maximally monotone, then  $R = \text{Id}$ .*

*Proof.* Suppose that  $(x, u) \in \text{gra } R$ , then we also have  $(u, x) \in \text{gra } R$ . Because of the monotonicity of  $R$  we now have  $0 \leq \langle x - u, u - x \rangle = -\|x - u\|^2 \leq 0$ , that is,  $x = u$ . ■

**Corollary 2.9** *The mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is maximally monotone if and only if  $\mathcal{R}(A, A^{-1}) = \text{Id}$ .*

*Proof.* If  $\mathcal{R}(A, A^{-1}) = \text{Id}$ , then since  $\text{Id}$  is maximally monotone, it is clear from Theorem 2.7 that  $A$  is maximally monotone. Conversely, suppose that the mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is maximally monotone. Then Theorem 2.2 guarantees that  $(\mathcal{R}(A, A^{-1}))^{-1} = \mathcal{R}(A^{-1}, A) = \mathcal{R}(A, A^{-1})$  while Theorem 2.7 guarantees that  $\mathcal{R}(A, A^{-1})$  is maximally monotone. Consequently, Proposition 2.8 implies that  $\mathcal{R}(A, A^{-1}) = \text{Id}$ . ■

Now we address domain and range properties of the resolvent average. A natural question is whether more precise Formulae than Formulae (32) can be attained. A precise formula for the domain of the proximal average of functions is [6, Theorem 4.6]. It asserts that given proper, convex and lower semicontinuous functions  $f_i : \mathcal{H} \rightarrow ]-\infty, +\infty]$ ,  $i \in I$ , then

$$\text{dom } p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = \sum_{i \in I} \lambda_i \text{dom } f_i.$$

As we shall see shortly, such precision does not hold in general for the resolvent average. However, we now aim at obtaining *nearly* precise formulae for the domain and range of the resolvent average. To this end, we first recall that the *relative interior* of a subset  $C$  of  $\mathcal{H}$ , which is denoted by  $\text{ri } C$ , is the subset of  $\mathcal{H}$  which is obtained by taking the interior of  $C$  when considered a subset of its closed affine hull. We say that the two subsets  $C$  and  $D$  of  $\mathcal{H}$  are *nearly equal* if  $\overline{C} = \overline{D}$  and  $\text{ri } C = \text{ri } D$ . In this case we write  $C \simeq D$ . In order to prove our near precise domain and range formulae we will employ the following fact which is an extension (to the case where we sum arbitrary finitely many rectangular mappings) of the classical and far reaching result of Brezis and Haraux [17] regarding the range of the sum of two rectangular mappings:

**Fact 2.10 (Pennanen)** [28, Corollary 6] *Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is monotone and rectangular. If  $A = \sum_{i \in I} A_i$  is maximally monotone, then*

$$\text{ran} \sum_{i \in I} A_i \simeq \sum_{i \in I} \text{ran} A_i. \quad (35)$$

**Theorem 2.11 (domain and range of  $\mathcal{R}_\mu$ )** *Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is maximally monotone. Then*

$$\text{ran } \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) \simeq \sum_{i \in I} \lambda_i \text{ran } A_i, \quad \text{dom } \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) \simeq \sum_{i \in I} \lambda_i \text{dom } A_i. \quad (36)$$

*Proof.* We first recall that a firmly nonexpansive mapping  $T : \mathcal{H} \rightarrow \mathcal{H}$  is maximally monotone (see [5, Example 20.27]) and rectangular (see [5, Example 24.16]). Thus, we see that all of the mappings  $J_{\mu A_i}$ ,  $i \in I$  as well as the firmly nonexpansive mapping  $J_{\mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})} = \sum_{i \in I} \lambda_i J_{\mu A_i}$  are maximally monotone and rectangular. As a consequence, we may now apply Fact 2.10 in order to obtain

$$\text{ran } J_{\mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})} \simeq \sum_{i \in I} \text{ran } \lambda_i J_{\mu A_i} = \sum_{i \in I} \lambda_i \text{ran } J_{\mu A_i}.$$

Since, given a maximally monotone mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$ , we have  $\text{dom } A = \text{dom } \mu A = \text{ran } J_{\mu A}$  (see Minty's parametrization (11)), we arrive at the domain formula in (36). We now combine the domain near equality with Theorem 2.2 in order to obtain

$$\text{ran } \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) = \text{dom}(\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}))^{-1} = \text{dom } \mathcal{R}_{\mu^{-1}}(\mathbf{A}^{-1}, \boldsymbol{\lambda}) \simeq \sum_{i \in I} \lambda_i \text{dom } A_i^{-1} = \sum_{i \in I} \lambda_i \text{ran } A_i.$$

■

We see that by employing the resolvent average we avoid constraint qualifications we had while employing the arithmetic average, one of the most obvious of which is we can now average mappings the domains of which do not intersect.

At this point we demonstrate why even in the finite-dimensional case the near equality in (36) cannot be replaced by an equality. To this end we consider the following example.

**Example 2.12** [2, Example 5.4] discussed the function  $f : \mathbb{R}^2 \rightarrow ]-\infty, +\infty]$ , defined by

$$f(x, y) = \begin{cases} -\sqrt{xy}, & x \geq 0, y \geq 0; \\ +\infty, & \text{otherwise.} \end{cases}$$

In [2] the function  $f$  was presented as an example of a proper, lower semicontinuous and sublinear function which is not subdifferentiable at certain points of the boundary of its domain, where the boundary is  $(\mathbb{R}_+ \times \{0\}) \cup (\{0\} \times \mathbb{R}_+)$ . In fact, it was observed in [2] that the only point on the boundary of  $\text{dom } \partial f$  which belongs to  $\text{dom } \partial f$  is the origin. We now consider the function  $g : \mathbb{R}^2 \rightarrow ]-\infty, +\infty]$  defined by

$$g(x, y) = \max\{f(1-x, y), f(1+x, y)\} = \begin{cases} -\sqrt{(1-|x|)y}, & -1 \leq x \leq 1, 0 \leq y; \\ +\infty, & \text{otherwise.} \end{cases}$$

We set  $D = \text{int dom } g = \{(x, y) \mid -1 < x < 1, 0 < y\}$ . Then it follows that  $g$  is lower semicontinuous, convex and  $\text{dom } \partial g = D \cup \{(-1, 0), (1, 0)\}$ . Now, we set  $n = 2$ ,  $A_1 = A_2 = \partial g$ ,  $\mu = 1$ ,  $0 < \lambda < 1$ ,  $\lambda_1 = \lambda$  and  $\lambda_2 = 1 - \lambda$ , then  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}) = \partial g$  (see formula (27)) and hence

$$\begin{aligned} \text{dom } \mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}) &= \text{dom } \partial g = D \cup \{(-1, 0), (1, 0)\} \\ &\neq D \cup \{(-1, 0), (1 - 2\lambda, 0), (2\lambda - 1, 0), (1, 0)\} = \lambda_1 \text{dom } A_1 + \lambda_2 \text{dom } A_2. \end{aligned}$$

Letting  $A_1 = A_2 = \partial g^* = (\partial g)^{-1}$  yields the same inequality with ranges instead of domains. Finally, we note that equality fails already in (35), that is, since  $\partial g^*$  is the subdifferential of a proper, lower semicontinuous and convex function, it is rectangular (see [5, Example 24.9]), maximally monotone and we have

$$\text{ran}(\partial g^* + \partial g^*) = 2D \cup \{(-2, 0), (2, 0)\} \neq 2D \cup \{(-2, 0), (0, 0), (2, 0)\} = \text{ran } \partial g^* + \text{ran } \partial g^*.$$

(For another example of this type, see [7, Example 3.14]).

## 2.4 The Fitzpatrick function of the resolvent average

We relate the Fitzpatrick function of the resolvent average with the Fitzpatrick functions of the averaged mappings in the following result:

**Theorem 2.13 (Fitzpatrick function of  $\mathcal{R}_\mu$ )** *Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is maximally monotone. Then*

$$F_{\mu\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})} \leq \sum_{i \in I} \lambda_i F_{\mu A_i} \quad \text{in particular,} \quad F_{\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})} \leq \sum_{i \in I} \lambda_i F_{A_i} \quad (37)$$

and

$$\sum_{i \in I} \lambda_i \text{dom } F_{\mu A_i} \subseteq \text{dom } F_{\mu\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})} \quad \text{in particular,} \quad \sum_{i \in I} \lambda_i \text{dom } F_{A_i} \subseteq \text{dom } F_{\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})}. \quad (38)$$

*Proof.* For each  $i \in I$ , let  $x_i, u_i$  and  $z$  be points in  $\mathcal{H}$  and let  $T_i = J_{\mu A_i}$ . We set  $(x, u) = \sum_{i \in I} \lambda_i (x_i, u_i)$  and  $T = \sum_{i \in I} \lambda_i T_i = J_{\mu\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})}$ . Then we employ equation (9) in order to obtain

$$\begin{aligned} \langle z - Tz, x \rangle + \langle u, Tz \rangle - \langle z - Tz, Tz \rangle &= \sum_{i \in I} \lambda_i (\langle z - T_i z, x_i \rangle + \langle u_i, T_i z \rangle - \langle z - T_i z, T_i z \rangle) \\ &\quad - \sum_{(i,j) \in I \times I} \frac{\lambda_i \lambda_j}{2} (\langle T_j z - T_i z, x_i - x_j \rangle + \langle u_i - u_j, T_i z - T_j z \rangle - \langle T_j z - T_i z, T_i z - T_j z \rangle) \\ &= \sum_{i \in I} \lambda_i (\langle z - T_i z, x_i \rangle + \langle u_i, T_i z \rangle - \langle z - T_i z, T_i z \rangle) \\ &\quad - \sum_{(i,j) \in I \times I} \frac{\lambda_i \lambda_j}{2} \left( -\frac{1}{4} \|(u_i - u_j) - (x_i - x_j)\|^2 + \left\| \frac{(u_i - u_j) - (x_i - x_j)}{2} + (T_i z - T_j z) \right\|^2 \right) \\ &\leq \sum_{i \in I} \lambda_i (\langle z - T_i z, x_i \rangle + \langle u_i, T_i z \rangle - \langle z - T_i z, T_i z \rangle) + \sum_{(i,j) \in I \times I} \frac{\lambda_i \lambda_j}{8} \|(u_i - u_j) - (x_i - x_j)\|^2. \quad (39) \end{aligned}$$

Given a maximally monotone mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$ , combining the definition (12) of the Fitzpatrick function  $F_A$  with Minty's parametrization (11) implies that

$$F_A(x, u) = \sup_{z \in \mathcal{H}} (\langle z - J_A z, x \rangle + \langle u, J_A z \rangle - \langle z - J_A z, J_A z \rangle).$$

Thus, by employing inequality (39) we obtain

$$F_{\mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})}(x, u) \leq \sum_{i \in I} \lambda_i F_{\mu A_i}(x_i, u_i) + \sum_{(i, j) \in I \times I} \frac{\lambda_i \lambda_j}{8} \|(u_i - u_j) - (x_i - x_j)\|^2. \quad (40)$$

For each  $i \in I$ , letting  $(x_i, u_i) = (x, u)$  in inequality (40) we arrive at inequality (37). For each  $i \in I$ , letting  $(x_i, u_i) \in \text{dom } F_{\mu A_i}$  in inequality (40) we see that  $F_{\mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})}(x, u) < \infty$ , that is, we obtain inclusions (38) and complete the proof.  $\blacksquare$

### 3 Dominant properties of the resolvent average

#### 3.1 Domain and range properties

The following domain and range properties of the resolvent average are immediate consequences of Theorem 2.11:

**Theorem 3.1 (nonempty interior of the domain, fullness of the domain and surjectivity are dominant)** *Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is maximally monotone.*

- (i) *If there exists  $i_0 \in I$  such that  $\text{int dom } A_{i_0} \neq \emptyset$ , then  $\text{int dom } \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) \neq \emptyset$ .*
- (ii) *If there exists  $i_0 \in I$  such that  $\text{dom } A_{i_0} = \mathcal{H}$ , then  $\text{dom } \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) = \mathcal{H}$ .*
- (iii) *If there exists  $i_0 \in I$  such that  $A_{i_0}$  is surjective, then  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is surjective.*

**Corollary 3.2 (nonempty interior and fullness of domain are dominant w.r.t.  $p_\mu$ )** *Suppose that for each  $i \in I$ ,  $f_i : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is proper, lower semicontinuous and convex. If there exists  $i_0 \in I$  such that  $\text{int dom } f_{i_0} \neq \emptyset$ , then  $\text{int dom } p_\mu(\mathbf{f}, \boldsymbol{\lambda}) \neq \emptyset$ . If there exists  $i_0 \in I$  such that  $\text{dom } f_{i_0} = \mathcal{H}$ , then  $\text{dom } p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = \mathcal{H}$ .*

*Proof.* For each  $i \in I$ , we set  $A_i = \partial f_i$ . Suppose that for some  $i_0 \in I$ ,  $\text{int dom } f_{i_0} \neq \emptyset$ . Since  $\text{int dom } f_{i_0} \subseteq \text{dom } \partial f_{i_0}$ , we see that  $\text{int dom } A_{i_0} \neq \emptyset$  and it now follows from Theorem 3.1(i) that  $\text{int dom } \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) \neq \emptyset$ . Now equation (7) implies that  $\emptyset \neq \text{int dom } \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) = \text{int dom } \partial p_\mu(\mathbf{f}, \boldsymbol{\lambda}) \subseteq \text{int dom } p_\mu(\mathbf{f}, \boldsymbol{\lambda})$ . A similar argument implies that if  $\text{dom } f_{i_0} = \mathcal{H}$ , then  $\text{dom } p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = \mathcal{H}$ .  $\blacksquare$

We recall that the function  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is said to be *coercive* if  $\lim_{\|x\| \rightarrow \infty} f(x) = \infty$ . The function  $f$  is said to be *supercoercive* if  $f/\|\cdot\|$  is coercive. As a consequence of Theorem 3.1(iii) in finite-dimensional spaces we obtain the following result:

**Corollary 3.3 (supercoercivity is dominant w.r.t.  $p_\mu$  in  $\mathbb{R}^n$ )** [22, Lemma 3.1(iii)] *Suppose that  $\mathcal{H}$  is finite-dimensional and that for each  $i \in I$ ,  $f_i : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is proper, lower semicontinuous and convex. If there exists  $i_0 \in I$  such that  $f_{i_0}$  is supercoercive, then  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is supercoercive.*

*Proof.* For each  $i \in I$ , we set  $A_i = \partial f_i$ . We now recall that in finite-dimensional spaces, a proper, lower semicontinuous and convex function  $f$  is supercoercive if and only if  $\text{dom } f^* = \mathcal{H}$ , which is equivalent to  $\mathcal{H} = \text{dom } \partial f^* = \text{ran } \partial f$  (combine [32, Corollary 13.3.1] with [32, Corollary 14.2.2], or, alternatively, see [3, Proposition 2.16]). Thus, since  $\text{ran } A_{i_0} = \mathcal{H}$ , then Theorem 3.1(iii) together with equation (7) guarantee that  $\text{ran } \partial p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = \text{ran } \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) = \mathcal{H}$ . Consequently, we see that  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is supercoercive.  $\blacksquare$

### 3.2 Single-valuedness

We shall say that the mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is at most single-valued if for every  $x \in \mathcal{H}$ ,  $Ax$  is either empty or a singleton.

**Theorem 3.4 (single-valuedness is dominant)** *Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is maximally monotone. If there exists  $i_0 \in I$  such that  $A_{i_0}$  is at most single-valued, then  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is at most single-valued.*

*Proof.* We recall that a maximally monotone mapping is at most single-valued if and only if its resolvent is injective (see [11, Theorem 2.1(iv)]). Thus, if  $A_{i_0}$  is at most single-valued, then  $\mu A_{i_0}$  is at most single-valued and  $J_{\mu A_{i_0}}$  is injective. We now apply Corollary 1.10(i) in order to conclude that  $J_{\mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})} = \sum_{i \in I} \lambda_i J_{\mu A_i}$  is injective and, consequently, that  $\mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is at most single-valued and so is  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$ .  $\blacksquare$

Recall that the proper, lower semicontinuous and convex function  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is said to be *essentially smooth* if the interior of its domain is nonempty and if  $\partial f$  is at most single-valued. This definition of essential smoothness coincides with the classical notions of the same name in finite-dimensional spaces (see the paragraph below preceding Corollary 3.8). In terms of essential smoothness of the proximal average, we recover the following result:

**Corollary 3.5 (essential smoothness is dominant w.r.t.  $p_\mu$ )** [6, Corollary 7.7] *Suppose that for each  $i \in I$ ,  $f_i : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is proper, lower semicontinuous and convex. If there exists  $i_0 \in I$  such that  $f_{i_0}$  is essentially smooth, then  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is essentially smooth.*

*Proof.* For each  $i \in I$ , we set  $A_i = \partial f_i$ . Suppose that for some  $i_0 \in I$ ,  $f_{i_0}$  is essentially smooth, then  $A_{i_0} = \partial f_{i_0}$  is at most single-valued and  $\text{int dom } A_{i_0} \neq \emptyset$ . It follows from Corollary 3.2 that  $\text{int dom } p_\mu(\mathbf{f}, \boldsymbol{\lambda}) \neq \emptyset$ . Furthermore, Theorem 3.4 guarantees that  $\partial p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is at most single-valued. Consequently,  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is essentially smooth.  $\blacksquare$



### 3.3 Strict monotonicity

Recall that  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is said to be *strictly monotone* if whenever  $u \in Ax$  and  $v \in Ay$  are such that  $x \neq y$ , then  $0 < \langle u - v, x - y \rangle$ . The resolvent perspective of this issue is addressed in the following proposition.

**Proposition 3.6** *Suppose that for each  $i \in I$ ,  $T_i : \mathcal{H} \rightarrow \mathcal{H}$  is firmly nonexpansive and set  $T = \sum_{i \in I} \lambda_i T_i$ . If there exists  $i_0 \in I$  such that*

$$T_{i_0}x \neq T_{i_0}y \quad \Rightarrow \quad \|T_{i_0}x - T_{i_0}y\|^2 < \langle x - y, T_{i_0}x - T_{i_0}y \rangle, \quad (41)$$

*then  $T$  has property (41) as well.*

*Proof.* Suppose that  $x$  and  $y$  are points in  $\mathcal{H}$  such that  $\|Tx - Ty\|^2 = \langle x - y, Tx - Ty \rangle$ . Then Corollary 1.10 implies that  $Tx - Ty = T_{i_0}x - T_{i_0}y$ . In particular, we see that

$$\|T_{i_0}x - T_{i_0}y\|^2 = \|Tx - Ty\|^2 = \langle x - y, Tx - Ty \rangle = \langle x - y, T_{i_0}x - T_{i_0}y \rangle.$$

Consequently, since  $T_{i_0}$  has property (41), we see that  $Tx - Ty = T_{i_0}x - T_{i_0}y = 0$ , as claimed.  $\blacksquare$

**Theorem 3.7 (strict monotonicity is dominant)** *Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is maximally monotone. If there exists  $i_0 \in I$  such that  $A_{i_0}$  is strictly monotone, then  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is strictly monotone.*

*Proof.* We recall that the maximally monotone mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is strictly monotone if and only if  $J_A$  has property (41) (see [11, Theorem 2.1(vi)]). Thus, since  $J_{\mu A_{i_0}}$  has property (41), then Proposition 3.6 guarantees that  $J_{\mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})} = \sum_{i \in I} \lambda_i J_{\mu A_i}$  has property (41), which, in turn, implies that  $\mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is strictly monotone and, therefore, so is  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$ .  $\blacksquare$

Recall that the proper, lower semicontinuous and convex function  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is said to be *essentially strictly convex* if  $f^*$  is essentially smooth;  $f$  is said to be *Legendre* if  $f$  is both, essentially smooth and essentially strictly convex. These definitions of essential strict convexity and Legendreness coincide with the classical notions of the same names in finite-dimensional spaces (see the next paragraph). Now suppose that for each  $i \in I$ ,  $f_i : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is proper, lower semicontinuous and convex. We suppose further that there exists  $i_0 \in I$  such that  $f_{i_0}$  is essentially strictly convex. Then  $f_{i_0}^*$  is essentially smooth. Consequently, Corollary 3.5 guarantees that  $p_{\mu^{-1}}(\mathbf{f}^*, \boldsymbol{\lambda})$  is essentially smooth. Thus, according to formula (15), the function  $p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = (p_\mu(\mathbf{f}, \boldsymbol{\lambda}))^{**} = (p_{\mu^{-1}}(\mathbf{f}^*, \boldsymbol{\lambda}))^*$  is essentially convex. Consequently, we see that essential strict convexity is dominant w.r.t. the proximal average. This line of proof of this fact was carried out in [6]. Since in the present paper formula (15), up to an additive constant, was recovered by Theorem 2.2, and since essential strict convexity is not affected by the addition of a constant to the function, we see that our discussion here does, indeed, recover the dominance of essential strict convexity w.r.t. the proximal average.

Classically, when  $\mathcal{H}$  is finite-dimensional, a different path that leads to the same conclusion is now available. Indeed, in this case, recall that the proper, lower semicontinuous and convex function

$f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is said to be essentially smooth if the interior of its domain is nonempty, if  $f$  is Gâteaux differentiable there and the norm of its Gâteaux gradient blows up as we tend from the interior to a point on the boundary of its domain.  $f$  is said to be essentially strictly convex if  $f$  is strictly convex on every convex subset of  $\text{dom } \partial f$ , which is equivalent to  $\partial f$  being strictly monotone (see [33, Theorem 12.17]). Thus, given that  $f_{i_0}$  is essentially strictly convex, then  $\partial f_{i_0}$  is strictly monotone. Setting  $A_i = \partial f_i$  for every  $i \in I$ , Theorem 3.7 guarantees that  $\partial p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is strictly monotone and, consequently, that  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is essentially strictly convex, as asserted. Summing up both of these discussions, we have recovered the following result:

**Corollary 3.8 (essential strict convexity is dominant w.r.t.  $p_\mu$ )** [6, Corollary 7.8] *Suppose that for each  $i \in I$ ,  $f_i : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is proper, lower semicontinuous and convex. If there exists  $i_0 \in I$  such that  $f_{i_0}$  is essentially strictly convex, then  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is essentially strictly convex.*

Combining Corollary 3.5 together with Corollary 3.8, we recover the following result:

**Corollary 3.9 (Legendreness is dominant w.r.t.  $p_\mu$ )** [6, Corollary 7.9] *Suppose that for each  $i \in I$ ,  $f_i : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is proper, lower semicontinuous and convex. If there exists  $i_1 \in I$  such that  $f_{i_1}$  is essentially smooth and there exists  $i_2 \in I$  such that  $f_{i_2}$  is essentially strictly convex, then  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is Legendre. In particular, if there exists  $i_0 \in I$  such that  $f_{i_0}$  is Legendre, then  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is Legendre.*

### 3.4 Uniform monotonicity

We say that the mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is monotone with modulus  $\varphi : [0, \infty[ \rightarrow [0, \infty]$  if for every two points  $(x, u)$  and  $(y, v)$  in  $\text{gra } A$ ,

$$\varphi(\|x - y\|) \leq \langle u - v, x - y \rangle.$$

Clearly, if  $\text{gra } A \neq \emptyset$ , then  $\varphi(0) = 0$ . We recall that the mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is said to be *uniformly monotone* with modulus  $\varphi$ ,  $\varphi$ -uniformly monotone for short, if  $A$  is monotone with modulus  $\varphi$  and  $\varphi(t) = 0 \Leftrightarrow t = 0$ . We say that the mapping  $T : \mathcal{H} \rightarrow \mathcal{H}$  is firmly nonexpansive with modulus  $\varphi : [0, \infty[ \rightarrow [0, \infty]$  if for every pair of points  $x$  and  $y$  in  $\mathcal{H}$ ,

$$\|Tx - Ty\|^2 + \varphi(\|Tx - Ty\|) \leq \langle Tx - Ty, x - y \rangle.$$

We also recall that the mapping  $T : \mathcal{H} \rightarrow \mathcal{H}$  is said to be *uniformly firmly nonexpansive* with modulus  $\varphi$ ,  $\varphi$ -uniformly firmly nonexpansive for short, if  $T$  is firmly nonexpansive with modulus  $\varphi$  and  $\varphi(t) = 0 \Leftrightarrow t = 0$ . For the sake of convenience, we will identify a modulus  $\varphi : [0, \infty[ \rightarrow [0, \infty]$  with  $\varphi(|\cdot|)$ , its symmetric extension to  $\mathbb{R}$ . With this convention, when we say that  $\varphi$  is increasing we mean that  $0 \leq t_1 < t_2 \Rightarrow \varphi(t_1) \leq \varphi(t_2)$ .

**Proposition 3.10** *Suppose that for each  $i \in I$ ,  $f_i : \mathcal{H} \rightarrow [0, \infty]$  is proper, lower semicontinuous, convex and  $f_i(0) = 0$ . Then  $p_\mu(\mathbf{f}, \boldsymbol{\lambda}) : \mathcal{H} \rightarrow [0, \infty]$  is proper, lower semicontinuous, convex and  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})(0) = 0$ . If there exists  $i_0 \in I$  such that  $f_{i_0}(x) = 0 \Leftrightarrow x = 0$ , then  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})(x) = 0 \Leftrightarrow x = 0$ .*

*Proof.* Fact 1.3 guarantees that  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is lower semicontinuous, convex and that for every  $x \in \mathcal{H}$ ,

$$p_\mu(\mathbf{f}, \boldsymbol{\lambda})(x) = \inf_{\sum_{i \in I} \lambda_i y_i = x} \sum_{i \in I} \lambda_i f_i(y_i) + \frac{1}{\mu} \left( \left( \sum_{i \in I} \lambda_i q(y_i) \right) - q(x) \right). \quad (42)$$

Since the bracketed term in (42) is greater or equal to zero and vanishes when for each  $i \in I$ ,  $y_i = x = 0$ , and since for each  $i \in I$ ,  $f_i$  is a function which is greater or equal to zero and which vanishes at zero, formula (42) implies that  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is greater or equal to zero and vanishes at zero. Finally, for each  $i \in I$ , we now set  $A_i = \partial f_i$ . Since for each  $i \in I$ , 0 is a minimizer of  $f_i$ , we see that  $\bigcap_{i \in I} A_i^{-1}(0) \neq \emptyset$ . If  $f_{i_0}(x) = 0 \Leftrightarrow x = 0$ , then  $\{0\} = A_{i_0}^{-1}(0) = \bigcap_{i \in I} A_i^{-1}(0)$ . Consequently, equation (30) together with equation (7) imply that  $\partial p_\mu(\mathbf{f}, \boldsymbol{\lambda})^{-1}(0) = \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})^{-1}(0) = \{0\}$ . We conclude that 0 is the only minimizer of  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  and, therefore,  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})(x) = 0 \Leftrightarrow x = 0$ . ■

**Proposition 3.11** *Suppose that for each  $i \in I$ ,  $T_i : \mathcal{H} \rightarrow \mathcal{H}$  is firmly nonexpansive with modulus  $\varphi_i$  which is lower semicontinuous and convex and set  $T = \sum_{i \in I} \lambda_i T_i$ . Then  $T$  is firmly nonexpansive with modulus  $\varphi = p_{\frac{1}{2}}(\boldsymbol{\varphi}, \boldsymbol{\lambda})$  which is proper, lower semicontinuous and convex. In particular, if there exists  $i_0 \in I$  such that  $T_{i_0}$  is  $\varphi_{i_0}$ -uniformly firmly nonexpansive, then  $T$  is  $\varphi$ -uniformly firmly nonexpansive.*

*Proof.* Fact 1.3 guarantees that  $\varphi = p_{\frac{1}{2}}(\boldsymbol{\varphi}, \boldsymbol{\lambda})$  is lower semicontinuous, convex and that for every  $t \in [0, \infty[$ ,

$$\varphi(t) = p_{\frac{1}{2}}(\boldsymbol{\varphi}, \boldsymbol{\lambda})(t) = \inf_{\sum_{i \in I} \lambda_i t_i = t} \sum_{i \in I} \lambda_i \varphi_i(t_i) + 2 \left( \left( \sum_{i \in I} \lambda_i q(t_i) \right) - q(t) \right). \quad (43)$$

Proposition 3.10 guarantees that  $\varphi$  is greater or equal to zero and vanishes at zero. Furthermore, if  $\varphi_{i_0}$  vanishes only at zero, then  $\varphi$  vanishes only at zero. Since  $\varphi$  is convex, we now see that it is increasing. Now, let  $x$  and  $y$  be points in  $\mathcal{H}$ . Then by employing formula (43) we obtain the following evaluation

$$\begin{aligned} \varphi(\|Tx - Ty\|) &\leq \varphi\left(\sum_{i \in I} \lambda_i \|T_i x - T_i y\|\right) \\ &\leq \sum_{i \in I} \lambda_i \varphi_i(\|T_i x - T_i y\|) + \sum_{i \in I} \lambda_i \|T_i x - T_i y\|^2 - \left(\sum_{i \in I} \lambda_i \|T_i x - T_i y\|\right)^2 \\ &\leq \sum_{i \in I} \lambda_i \varphi_i(\|T_i x - T_i y\|) + \sum_{i \in I} \lambda_i \|T_i x - T_i y\|^2 - \left\| \sum_{i \in I} \lambda_i (T_i x - T_i y) \right\|^2 \end{aligned}$$

which implies that

$$\begin{aligned} \varphi(\|Tx - Ty\|) + \|Tx - Ty\|^2 &\leq \sum_{i \in I} \lambda_i \left( \varphi_i(\|T_i x - T_i y\|) + \|T_i x - T_i y\|^2 \right) \\ &\leq \sum_{i \in I} \lambda_i \langle T_i x - T_i y, x - y \rangle = \langle Tx - Ty, x - y \rangle. \end{aligned}$$

Thus, we see that  $T$  is firmly nonexpansive with modulus  $\varphi$ . ■

As a consequence, we obtain the following result:

**Theorem 3.12 (uniform monotonicity with a convex modulus is dominant)** *Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is maximally monotone with modulus  $\varphi_i$  which is lower semicontinuous and convex. Then  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is monotone with modulus  $\varphi = p_{\frac{\mu}{2}}(\boldsymbol{\varphi}, \boldsymbol{\lambda})$  which is lower semicontinuous and convex. In particular, if there exists  $i_0 \in I$  such that  $A_{i_0}$  is  $\varphi_{i_0}$ -uniformly monotone, then  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is  $\varphi$ -uniformly monotone.*

*Proof.* First we consider the case  $\mu = 1$ . To this end we will employ the fact that the maximally monotone mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is monotone with modulus  $\varphi$  if and only if  $J_A$  is firmly nonexpansive with modulus  $\varphi$ . Indeed, employing Minty's parametrization (11), we see that  $A$  is uniformly monotone with modulus  $\varphi$  if and only if for every  $x$  and  $y$  in  $\mathcal{H}$  we have  $\varphi(\|J_A x - J_A y\|) \leq \langle J_A x - J_A y, (x - y) - (J_A x - J_A y) \rangle$  which is precisely the firm nonexpansiveness of  $J_A$  with modulus  $\varphi$ . Thus, we see that for each  $i \in I$ ,  $J_{A_i}$  is firmly nonexpansive with modulus  $\varphi_i$ . Consequently, Proposition 3.11 guarantees that  $J_{\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})} = \sum_{i \in I} \lambda_i J_{A_i}$  is firmly nonexpansive with modulus  $\varphi = p_{\frac{1}{2}}(\boldsymbol{\varphi}, \boldsymbol{\lambda})$ , which, in turn, implies that  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$  is monotone with modulus  $\varphi$ . For an arbitrary  $0 < \mu$ , we employ formulae (26) and (28) as follows: any mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is  $\varphi$ -monotone if and only if  $\mu A$  is  $\mu\varphi$ -monotone. Thus, since we already have that  $\mathcal{R}(\mu\mathbf{A}, \boldsymbol{\lambda})$  is  $p_{\frac{1}{2}}(\mu\boldsymbol{\varphi}, \boldsymbol{\lambda})$ -monotone, then  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) = \mu^{-1}\mathcal{R}(\mu\mathbf{A}, \boldsymbol{\lambda})$  is  $\varphi$ -monotone where  $\varphi = \mu^{-1}p_{\frac{1}{2}}(\mu\boldsymbol{\varphi}, \boldsymbol{\lambda}) = \frac{2}{\mu}p(\frac{\mu}{2}\boldsymbol{\varphi}, \boldsymbol{\lambda}) = p_{\frac{\mu}{2}}(\boldsymbol{\varphi}, \boldsymbol{\lambda})$ . Finally, Proposition 3.10 guarantees that if  $\varphi_{i_0}$  vanishes only at zero, then  $\varphi$  vanishes only at zero.  $\blacksquare$

We recall (see [37, Section 3.5]) that the proper function  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is said to be *uniformly convex* if there exists a function  $\varphi : [0, \infty[ \rightarrow [0, \infty]$  with the property  $\varphi(t) = 0 \Leftrightarrow t = 0$  such that for every two points  $x$  and  $y$  in  $\mathcal{H}$  and every  $\lambda \in ]0, 1[$ ,

$$f((1 - \lambda)x + \lambda y) + \lambda(1 - \lambda)\varphi(\|x - y\|) \leq (1 - \lambda)f(x) + \lambda f(y). \quad (44)$$

The largest possible function  $\varphi$  satisfying (44) is called the *gauge of uniform convexity* of  $f$  and is defined by

$$\varphi_f(t) = \inf \left\{ \frac{(1 - \lambda)f(x) + \lambda f(y) - f((1 - \lambda)x + \lambda y)}{\lambda(1 - \lambda)} \mid \lambda \in ]0, 1[, x, y \in \text{dom } f, \|x - y\| = t \right\}.$$

We also recall that the proper and convex function  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is said to be *uniformly smooth* if there exists a function  $\psi : [0, \infty[ \rightarrow [0, \infty]$  with the property  $\lim_{t \rightarrow 0} \psi(t)/t = 0$  such that for every two points  $x$  and  $y$  in  $\mathcal{H}$  and every  $\lambda \in ]0, 1[$  such that  $(1 - \lambda)x + \lambda y \in \text{dom } f$ ,

$$f((1 - \lambda)x + \lambda y) + \lambda(1 - \lambda)\psi(\|x - y\|) \geq (1 - \lambda)f(x) + \lambda f(y). \quad (45)$$

The smallest possible function  $\psi$  satisfying (45) is called the *gauge of uniform smoothness* of  $f$  and is defined by

$$\psi_f(t) = \sup \left\{ \frac{(1 - \lambda)f(x) + \lambda f(y) - f((1 - \lambda)x + \lambda y)}{\lambda(1 - \lambda)} \mid \lambda \in ]0, 1[, x, y \in \mathcal{H}, \|x - y\| = t, \right. \\ \left. ((1 - \lambda)x + \lambda y) \in \text{dom } f \right\}.$$

**Fact 3.13** [37, Theorem 3.5.10, (i) $\Leftrightarrow$ (v)] *Suppose that  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is proper, lower semicontinuous and convex. Then  $f$  is uniformly convex if and only if  $\partial f$  is uniformly monotone, in which case  $\partial f$  is  $\varphi$ -uniformly monotone with  $\varphi = 2\varphi_f^{**}$ .*

Consequently, we arrive at the following results:

**Theorem 3.14 (uniform convexity is dominant w.r.t.  $p_\mu$ )** *Suppose that for each  $i \in I$ ,  $f_i : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is proper, lower semicontinuous, convex and there exists  $i_0 \in I$  such that  $f_{i_0}$  is uniformly convex. Then  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is uniformly convex.*

*Proof.* For each  $i \in I$  we set  $A_i = \partial f_i$ . Since  $f_{i_0}$  is uniformly convex, Fact 3.13 guarantees that there exists a lower semicontinuous and convex modulus  $\varphi_{i_0}$  such that  $A_{i_0}$  is  $\varphi_{i_0}$ -uniformly monotone. Consequently, Theorem 3.12 together with equation (7) guarantee that  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) = \partial p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is uniformly monotone, which, in turn, implies that  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is uniformly convex.  $\blacksquare$

**Fact 3.15** [37, Theorem 3.5.5] *Suppose that  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is proper, lower semicontinuous and convex. Then (i)  $f$  is uniformly convex if and only if  $f^*$  is uniformly smooth and (ii)  $f$  is uniformly smooth if and only if  $f^*$  is uniformly convex. (Within our reflexive settings (i) and (ii) are equivalent.)*

**Theorem 3.16 (uniform smoothness is dominant w.r.t.  $p_\mu$ )** *Suppose that for each  $i \in I$ ,  $f_i : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is proper, lower semicontinuous, convex and there exists  $i_0 \in I$  such that  $f_{i_0}$  is uniformly smooth. Then  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is uniformly smooth.*

*Proof.* Fact 3.15 guarantees that  $f_{i_0}^*$  is uniformly convex. Consequently, by Theorem 3.14,  $p_{\mu^{-1}}(\mathbf{f}^*, \boldsymbol{\lambda})$  is uniformly convex. Finally, Applying Fact 3.15 together with formula (15) implies that  $p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = p_\mu(\mathbf{f}, \boldsymbol{\lambda})^{**} = p_{\mu^{-1}}(\mathbf{f}^*, \boldsymbol{\lambda})^*$  is uniformly smooth.  $\blacksquare$

**Remark 3.17** A remark regarding the sharpness of our results is now in order. Under the generality of the hypotheses of Theorem 3.12, even when given that for each  $i \in I$ ,  $\varphi_i$  is the largest possible modulus of monotonicity of  $A_i$ , it does not hold that  $\varphi = p_\mu^\mu(\boldsymbol{\varphi}, \boldsymbol{\lambda})$  is necessarily the largest possible modulus of monotonicity of  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$ . In fact, the largest modulus of monotonicity of  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  cannot, in general, be expressed only in terms of  $\mu, \boldsymbol{\varphi}$  and  $\boldsymbol{\lambda}$  but depends also on  $\mathbf{A}$ . To this end, we consider the following example which illustrates a similar situation for functions and, consequently, for their subdifferential operators. We will illustrate this issue outside the class of subdifferential operators in Example 3.29 of the following subsection.

**Example 3.18** Let  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  be a proper, lower semicontinuous and convex function such that  $\varphi_f = 0$  and  $\psi_f = \iota_{\{0\}}$ . Among a rich variety of choices, one can choose  $f = \iota_C$  where  $C$  is a closed half-space. For such a choice of  $f$ , one easily verifies that, indeed, by their definitions,  $\varphi_f = 0$  and  $\psi_f = \iota_{\{0\}}$ . Then it follows that  $f$  is neither uniformly convex nor uniformly smooth. Consequently, Fact 3.15 implies that also  $f^*$  is neither uniformly convex nor uniformly smooth. However, employing formula (34), we see that  $p(f, f^*) = q$  which is both, uniformly convex and uniformly smooth (in fact, as we recall in the following subsection,  $q$  is strongly convex and strongly smooth). Thus, we see that  $\varphi_{p_\mu(\mathbf{f}, \boldsymbol{\lambda})}$  and  $\psi_{p_\mu(\mathbf{f}, \boldsymbol{\lambda})}$  cannot, in general, be expressed only in terms of  $\mu, \boldsymbol{\varphi}_f, \boldsymbol{\psi}_f$  and  $\boldsymbol{\lambda}$  since, for example, in the case of  $p(f, f) = f$  we have the same weights  $\lambda_1 = \lambda_2 = \frac{1}{2}$ , the same  $\mu = 1$  and the averaged functions have the same gages of uniform convexity and smoothness which are 0 and  $\iota_{\{0\}}$ , respectively, as in the case  $p(f, f^*) = q$ . However,  $\varphi_{p(f, f)} = 0$  and  $\psi_{p(f, f)} = \iota_{\{0\}}$ , that is,  $p(f, f)$  is neither uniformly convex nor uniformly smooth.

**Remark 3.19** Before ending the current discussion a remark regarding the hypothesis of Theorem 3.12 is in order. In the general framework of monotone operators, we are not aware of a study of finer properties of the modulus of uniform monotonicity. In particular, existence of a convex modulus (like in the case of uniformly monotone subdifferential operators for which there exists a modulus which is convex and which possesses also other attractive properties (see [37, Section 3.5]), which is crucial in the proof of Theorem 3.12, is unavailable to us. At this point we relegate such a finer study for future research.

### 3.5 Strong monotonicity and cocoercivity

We say that the mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is  $\epsilon$ -monotone, where  $\epsilon \geq 0$ , if  $A - \epsilon \text{Id}$  is monotone, that is, if for any two points  $(x, u)$  and  $(y, v)$  in  $\text{gra } A$ ,

$$\epsilon \|x - y\|^2 \leq \langle v - u, x - y \rangle.$$

In particular, a 0-monotone mapping is simply a monotone mapping. We now recall that  $A$  is said to be strongly monotone with constant  $\epsilon$ ,  $\epsilon$ -strongly monotone for short, if it is  $\epsilon$ -monotone with  $0 < \epsilon$ . Clearly,  $\epsilon$ -monotone implies  $\epsilon'$ -monotone for any  $0 \leq \epsilon' < \epsilon$ . Letting  $\varphi(t) = \epsilon t^2$ , we see that  $A$  is  $\epsilon$ -monotone if and only if  $A$  is monotone with modulus  $\varphi$ . Thus, the subject matter of our current discussion is a particular case of our discussion in the proceeding subsection. However, in view of the importance of strong monotonicity, cocoercivity, strong convexity, strong smoothness and Lipschitzness of the gradient (all which will be defined shortly), we single out these subjects and treat them separately. Moreover, in the present discussion we add quantitative information in terms of explicit constants of the above properties. To this end, we will make use of the following notations and conventions.

First we fix the following conventions in  $[0, \infty]$ :  $0^{-1} = \infty$ ,  $\infty^{-1} = 0$  and  $0 \cdot \infty = 0$ . For  $S \subseteq \mathcal{H}$  we fix  $\infty \cdot S = \mathcal{H}$  if  $0 \in S$  and  $\infty \cdot S = \emptyset$  if  $0 \notin S$ . We will apply these conventions in the case where  $\mathcal{H} = \mathbb{R}$  in order to calculate terms of the form  $\alpha \text{Id}$  and  $\alpha q$  where  $\alpha \in [0, \infty]$ , however, these calculation hold in any Hilbert space  $\mathcal{H}$ . With these conventions at hand, it now follows that  $\infty q = \iota_{\{0\}}$ ,  $(\infty q)^* = 0$  and  $\partial(\infty q) = \partial \iota_{\{0\}} = N_{\{0\}} = \infty \cdot \text{Id} = \infty \partial q$ . Consequently, for every  $\alpha \in [0, \infty]$  we deduce the following formulae:  $(\alpha q)^* = \alpha^{-1} q$ ,  $\partial \alpha q = \alpha \text{Id} = \alpha \partial q$  and  $\partial(\alpha q)^* = \alpha^{-1} \text{Id} = (\alpha \text{Id})^{-1} = (\partial \alpha q)^{-1}$ .

Suppose that for each  $i \in I$ ,  $0 \leq \alpha_i \leq \infty$  and set  $\boldsymbol{\alpha} = (\alpha_1 \cdots, \alpha_n)$ . Then we define

$$r_\mu(\boldsymbol{\alpha}, \boldsymbol{\lambda}) = \left[ \sum_{i \in I} \lambda_i (\alpha_i + \mu^{-1})^{-1} \right]^{-1} - \mu^{-1} \quad \text{and} \quad r(\boldsymbol{\alpha}, \boldsymbol{\lambda}) = r_1(\boldsymbol{\alpha}, \boldsymbol{\lambda}). \quad (46)$$

We note that  $0 < r_\mu(\boldsymbol{\alpha}, \boldsymbol{\lambda})$  if and only if there exists  $i_0 \in I$  such that  $0 < \alpha_{i_0}$  and  $r_\mu(\boldsymbol{\alpha}, \boldsymbol{\lambda}) < \infty$  if and only if there exists  $i_0 \in I$  such that  $\alpha_{i_0} < \infty$ . For each  $i \in I$  we set  $f_i = \alpha_i q$  and  $A_i = \alpha_i \text{Id} = \partial f_i$ . Then by combining our settings and calculations above with either a direct computations or with our formulae from Section 2 (namely, formulae (4), (15), (20) and (26)), the



following properties of  $r_\mu$  follow:

$$\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) = r_\mu(\boldsymbol{\alpha}, \boldsymbol{\lambda}) \text{Id}, \quad (47)$$

$$p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = r_\mu(\boldsymbol{\alpha}, \boldsymbol{\lambda})q, \quad (48)$$

$$r_\mu(\boldsymbol{\alpha}, \boldsymbol{\lambda})^{-1} = r_{\mu^{-1}}(\boldsymbol{\alpha}^{-1}, \boldsymbol{\lambda}), \quad \text{that is,} \quad r_{\mu^{-1}}(\boldsymbol{\alpha}^{-1}, \boldsymbol{\lambda})^{-1} = r_\mu(\boldsymbol{\alpha}, \boldsymbol{\lambda}), \quad (49)$$

$$r_\mu(\boldsymbol{\alpha}, \boldsymbol{\lambda}) = \mu^{-1}r(\mu\boldsymbol{\alpha}, \boldsymbol{\lambda}). \quad (50)$$

For positive real numbers  $\alpha_i$ , formula (48) was obtained in [6, formula (25)].

**Theorem 3.20 (strong monotonicity is dominant)** *Suppose that for each  $i \in I$ ,  $\epsilon_i \geq 0$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is maximally monotone and  $\epsilon_i$ -monotone. Then  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is  $\epsilon$ -monotone where  $\epsilon = r_\mu(\boldsymbol{\epsilon}, \boldsymbol{\lambda})$ . In particular, if there exists  $i_0 \in I$ , such that  $A_{i_0}$  is  $\epsilon_{i_0}$ -strongly monotone, then  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is  $\epsilon$ -strongly monotone.*

*Proof.* For each  $i \in I$  we let  $\varphi_i(t) = \epsilon_i t^2 = 2\epsilon_i(t^2/2)$ . Then  $A_i$  is monotone with modulus  $\varphi_i$ . Consequently, Theorem 3.12 guarantees that  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is monotone with modulus  $\varphi = p_\mu(\boldsymbol{\varphi}, \boldsymbol{\lambda})$ . By employing formulae (48) and (50) we see that for every  $t \geq 0$ ,  $\varphi(t) = r_{\frac{\mu}{2}}(2\boldsymbol{\epsilon}, \boldsymbol{\lambda})(t^2/2) = r_\mu(\boldsymbol{\epsilon}, \boldsymbol{\lambda})t^2$  which means that  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is  $\epsilon$ -monotone where  $\epsilon = r_\mu(\boldsymbol{\epsilon}, \boldsymbol{\lambda})$ . In particular, if there exists  $i_0 \in I$ , such that  $\epsilon_{i_0} > 0$ , then  $r_\mu(\boldsymbol{\epsilon}, \boldsymbol{\lambda}) > 0$ . ■

We recall that the mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is said to be  $\epsilon$ -cocoercive, where  $\epsilon > 0$ , if  $A^{-1}$  is  $\epsilon$ -strongly monotone, that is, if for every pair of points  $(x, u)$  and  $(y, v)$  in  $\text{gra } A$ ,  $\epsilon\|u - v\|^2 \leq \langle u - v, x - y \rangle$ . The following result is an immediate consequence of Theorem 3.20.

**Corollary 3.21 (cocoerciveness is dominant)** *Suppose that for each  $i \in I$ ,  $\epsilon_i \geq 0$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is maximally monotone and  $A_i^{-1}$  is  $\epsilon_i$ -monotone. Then  $(\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}))^{-1}$  is  $\epsilon$ -monotone where  $\epsilon = r_{\mu^{-1}}(\boldsymbol{\epsilon}, \boldsymbol{\lambda})$ . In particular, if there exists  $i_0 \in I$  such that  $A_{i_0}$  is  $\epsilon_{i_0}$ -cocoercive, then  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$  is  $\epsilon$ -cocoercive.*

*Proof.* Since for each  $i \in I$ ,  $A_i^{-1}$  is  $\epsilon_i$ -monotone, then Theorem 3.20 together with Theorem 2.2 guarantee that  $(\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}))^{-1} = \mathcal{R}_{\mu^{-1}}(\mathbf{A}^{-1}, \boldsymbol{\lambda})$  is  $\epsilon$ -monotone. ■

We say that the proper function  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is  $\epsilon$ -convex, where  $\epsilon \geq 0$ , if  $f$  is convex with modulus  $\varphi(t) = \frac{\epsilon}{2}t^2$  (see (44)). In particular, a 0-convex function is simply a convex function. We recall that  $f$  is said to be *strongly convex* with constant  $\epsilon$ ,  $\epsilon$ -strongly convex for short, if  $f$  is  $\epsilon$ -convex and  $\epsilon > 0$ . We say that the proper and convex function  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is  $\epsilon$ -smooth, where  $\epsilon \geq 0$ , if  $f$  is smooth with modulus  $\psi(t) = \frac{1}{2\epsilon}t^2$  (see (45)). In particular, a 0-smooth function is any proper and convex function. We now recall that  $f$  is said to be *strongly smooth* with constant  $\epsilon$ ,  $\epsilon$ -strongly smooth for short, if  $f$  is  $\epsilon$ -smooth and  $\epsilon > 0$ .

**Fact 3.22** [37, Corollary 3.5.11 and Remark 3.5.3] *Suppose that  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is proper, lower semicontinuous and convex. Then the following assertions are equivalent:*

- (i)  $f$  is  $\epsilon$ -convex;

(ii)  $\partial f$  is  $\epsilon$ -monotone;

(iii)  $f^*$  is  $\epsilon$ -smooth.

If  $\epsilon > 0$ , assertions (i), (ii) and (iii) above are equivalent to the following assertion:

(iv)  $\text{dom } f^* = \mathcal{H}$ ,  $f^*$  is Fréchet differentiable on  $\mathcal{H}$  and  $\partial f^* = \nabla f^*$  is  $\frac{1}{\epsilon}$ -Lipschitz.

As a consequence, we obtain the following results:

**Theorem 3.23 (strong convexity is dominant w.r.t.  $p_\mu$ )** Suppose that for each  $i \in I$ ,  $f_i : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is proper, lower semicontinuous and  $\epsilon_i$ -convex. Then  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is  $\epsilon$ -convex where  $\epsilon = r_\mu(\boldsymbol{\epsilon}, \boldsymbol{\lambda})$ . In particular, if there exists  $i_0 \in I$  such that  $f_{i_0}$  is  $\epsilon_{i_0}$ -strongly convex, then  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is  $\epsilon$ -strongly convex.

*Proof.* For each  $i \in I$  we set  $A_i = \partial f_i$ . Then Fact 3.22 guarantees that  $A_i$  is  $\epsilon_i$ -monotone. Consequently, Theorem 3.20 together with equation (7) imply that  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) = \partial p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is  $\epsilon$ -monotone which, in turn, implies that  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is  $\epsilon$ -convex. In particular, if there exists  $i_0 \in I$  such that  $f_{i_0}$  is  $\epsilon_{i_0}$ -strongly convex, then  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is  $\epsilon$ -strongly convex. ■

**Theorem 3.24 (strong smoothness is dominant w.r.t.  $p_\mu$ )** Suppose that for each  $i \in I$ ,  $f_i : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is proper, lower semicontinuous and  $\epsilon_i$ -smooth. Then  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is  $\epsilon$ -smooth where  $\epsilon = r_{\mu^{-1}}(\boldsymbol{\epsilon}, \boldsymbol{\lambda})$ . In particular, if there exists  $i_0 \in I$  such that  $f_{i_0}$  is  $\epsilon_{i_0}$ -strongly smooth, then  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is  $\epsilon$ -strongly smooth.

*Proof.* Fact 3.22 asserts that each  $f_i^*$  is  $\epsilon$ -convex. Consequently, Theorem 3.23 and formula (15) guaranty that  $p_{\mu^{-1}}(\mathbf{f}^*, \boldsymbol{\lambda}) = p_\mu(\mathbf{f}, \boldsymbol{\lambda})^*$  is  $\epsilon$ -convex which, in turn, imply that  $p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = p_\mu(\mathbf{f}, \boldsymbol{\lambda})^{**} = p_{\mu^{-1}}(\mathbf{f}^*, \boldsymbol{\lambda})^*$  is  $\epsilon$ -smooth. ■

**Theorem 3.25 (having a Lipschitz gradient is dominant w.r.t.  $p_\mu$ )** Suppose that for each  $i \in I$ ,  $f_i : \mathcal{H} \rightarrow ]-\infty, +\infty]$  is proper, lower semicontinuous, convex and set  $A_i = \partial f_i$ . Suppose further that there exist  $\emptyset \neq I_0 \subseteq I$  such that for every  $i \in I_0$ ,  $f_i$  is Fréchet differentiable on  $\mathcal{H}$ ,  $\nabla f_i$  is  $\epsilon_i$ -Lipschitz and for every  $i \notin I_0$  set  $\epsilon_i = \infty$ . Then  $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is Fréchet differentiable on  $\mathcal{H}$  and  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) = \nabla p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is  $\epsilon$ -Lipschitz where  $\epsilon = r_\mu(\boldsymbol{\epsilon}, \boldsymbol{\lambda})$ .

*Proof.* Fact 3.22 guarantees that for each  $i \in I$ ,  $f_i^*$  is  $\frac{1}{\epsilon_i}$ -convex. By applying Theorem 3.23 we see that  $p_{\mu^{-1}}(\mathbf{f}^*, \boldsymbol{\lambda})$  is  $\frac{1}{\epsilon}$ -convex where  $\frac{1}{\epsilon} = r_{\mu^{-1}}(\boldsymbol{\epsilon}^{-1}, \boldsymbol{\lambda})$ . Furthermore, since  $I_0 \neq \emptyset$ , we see that  $0 < \frac{1}{\epsilon} < \infty$  and, consequently, that  $p_{\mu^{-1}}(\mathbf{f}^*, \boldsymbol{\lambda})$  is  $\frac{1}{\epsilon}$ -strongly convex. By applying Fact 3.22 together with equation (7) and equation (15) we conclude that  $p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = p_{\mu^{-1}}(\mathbf{f}^*, \boldsymbol{\lambda})^*$  is Fréchet differentiable on  $\mathcal{H}$  and that  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) = \partial p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = \nabla p_\mu(\mathbf{f}, \boldsymbol{\lambda})$  is  $\epsilon$ -Lipschitz where, by applying formula (49),  $\epsilon = r_{\mu^{-1}}(\boldsymbol{\epsilon}^{-1}, \boldsymbol{\lambda})^{-1} = r_\mu(\boldsymbol{\epsilon}, \boldsymbol{\lambda})$ . ■



**Remark 3.26** A remark regarding the sharpness of our results is now in order. Although in particular cases, our constants of monotonicity, Lipschitzness and cocoercivity of the resolvent average as well as the constants of convexity and smoothness of the proximal average are sharp, in general, they are not. Furthermore, such sharp constants cannot be determined only by  $\mu$ , the weight  $\lambda$  and the given constant  $\epsilon$  even if the latter is sharp, but depend also on the averaged objects as well. We now illustrate these situations by the following examples.

**Example 3.27** For each  $i \in I$  let  $0 < \epsilon_i < \infty$ ,  $f_i = \epsilon_i q$  and  $A_i = \epsilon_i \text{Id} = \nabla f_i$ . Then for each  $i \in I$ ,  $A_i$  is  $\epsilon_i$ -strongly monotone which is equivalent to  $f_i$  being  $\epsilon_i$ -strongly convex. As we have seen by formulae (47) and (48) we have  $\mathcal{R}_\mu(\mathbf{A}, \lambda) = r_\mu(\epsilon, \lambda) \text{Id}$  which is  $r_\mu(\epsilon, \lambda)$ -strongly monotone and  $p_\mu(\mathbf{f}, \lambda) = r_\mu(\epsilon, \lambda)q$  which is  $r_\mu(\epsilon, \lambda)$ -strongly convex. Furthermore, for any  $\epsilon' > r_\mu(\epsilon, \lambda)$  we see that  $\mathcal{R}_\mu(\mathbf{A}, \lambda) - \epsilon' \text{Id}$  is not monotone, that is,  $\mathcal{R}_\mu(\mathbf{A}, \lambda)$  is not  $\epsilon'$ -monotone, and that  $p_\mu(\mathbf{f}, \lambda) - \epsilon'q$  is not convex, that is,  $p_\mu(\mathbf{f}, \lambda)$  is not  $\epsilon'$ -convex. Thus, we conclude that  $\epsilon = r_\mu(\epsilon, \lambda)$  is a sharp constant of strong monotonicity of the resolvent average and as a constant of strong convexity of the proximal average in this example.

**Example 3.28** We consider the settings in Example 3.18. We see that neither  $f$  nor  $f^*$  is strongly convex nor strongly smooth. Also, neither  $\partial f$  nor  $\partial f^*$  is strongly monotone or Lipschitz continuous. However,  $p(f, f^*) = q$  is 1-strongly convex and 1-strongly smooth and  $\mathcal{R}(\partial f, \partial f^*) = \text{Id}$  is 1-strongly monotone and 1-Lipschitz. Letting  $\epsilon_1 = \epsilon_2 = 0$  (which is the only possible constant of monotonicity of  $\partial f$  and of  $\partial f^*$ , and the only possible constant of convexity and smoothness of  $f$  and  $f^*$ ), we see that  $\epsilon = r(\epsilon, \lambda) = 0$ . Thus, we conclude that  $\epsilon$  is not sharp in this case. On the other hand,  $\epsilon = 0$  will be the sharp constant of convexity when we consider  $p(0, 0) = 0$  and a sharp constant of monotonicity and Lipschitzness of  $\mathcal{R}(0, 0) = 0$  where the given  $\mu$ ,  $\lambda$  and  $\epsilon$  are the same. Thus we conclude that the sharp constants of monotonicity, Lipschitzness and cocoercivity of the resolvent average as well as the sharp constants of convexity and smoothness of the proximal average cannot, in general, be determined only by the corresponding constants of the averaged objects, the parameter  $\mu$  and the weight  $\lambda$ .

**Example 3.29** Outside the class of subdifferential operators, we consider the following settings: let  $\lambda_1 = \lambda_2 = \frac{1}{2}$ , let  $A_1$  be the counterclockwise rotation in the plane by the angle  $0 < \alpha \leq \frac{\pi}{2}$  and let  $A_2 = A_1^{-1}$  be the clockwise rotation by the angle  $\alpha$ . Then  $A_1$  and  $A_2$  are  $\epsilon$ -monotone where  $\epsilon = \cos \alpha < 1$  is sharp. In particular, for  $\alpha = \frac{\pi}{2}$ ,  $A_1$  and  $A_2$  are not strongly monotone. However, by employing formula (2.8), we see that  $\mathcal{R}(A_1, A_2) = \text{Id}$  which is 1-strongly monotone. On the other hand,  $\mathcal{R}(A_1, A_1) = A_1$  is  $\epsilon$ -monotone where, again,  $\epsilon = \cos \alpha$  is sharp. Thus, this is another demonstration to the fact that the sharp constant of monotonicity of the resolvent average does not depend only on the constants of monotonicity of the averaged mappings, the weight  $\lambda$  and the parameter  $\mu$ . We will analyze a variant of this example in order to discuss Lipschitzness outside the class of subdifferential operators in Example 5.11.

### 3.6 Disjoint injectivity

We recall that the mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is said to be *disjointly injective* if for any two distinct point  $x$  and  $y$  in  $\mathcal{H}$ ,  $Ax \cap Ay = \emptyset$ .

**Theorem 3.30 (disjoint injectivity is dominant)** *Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is maximally monotone. If there exists  $i_0 \in I$  such that  $A_{i_0}$  is disjointly injective, then  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is disjointly injective.*

*Proof.* We recall that the maximally monotone mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is disjointly injective if and only if  $J_A$  is strictly nonexpansive, that is, for any two distinct points  $x$  and  $y$  in  $\mathcal{H}$ ,  $\|J_A x - J_A y\| < \|x - y\|$  (see [11, Theorem 2.1(ix)]). We also note that  $A$  is disjointly injective if and only if  $\mu A$  is disjointly injective. Thus, since for every  $i \in I$ ,  $J_{\mu A_i}$  is nonexpansive and since  $J_{\mu A_{i_0}}$  is strictly nonexpansive, then the convex combination  $J_{\mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})} = \sum_{i \in I} \lambda_i J_{\mu A_i}$  is strictly nonexpansive, which, in turn, implies that  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is disjointly injective.  $\blacksquare$

**Remark 3.31** Let  $\mathcal{H}$  be finite-dimensional and let  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  be proper, lower semicontinuous and convex function. Then the disjoint injectivity of  $\partial f$  is equivalent to the essential strict convexity of  $f$  which is equivalent to the essential smoothness of  $f^*$  (see [32, Theorem 26.3]). Thus, in the finite-dimensional case, once again, we recover Corollary 3.5, Corollary 3.8 and Corollary 3.9, that is, essential smoothness, essential strict convexity and Legendreness are dominant properties w.r.t. the proximal average.

## 4 Recessive properties of the resolvent average

We begin this section with the following example of a recessive property:

**Example 4.1 (being a constant mapping is recessive)** Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is the constant mapping  $x \mapsto z_i$ . Then  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is the constant mapping  $x \mapsto \sum_{i \in I} \lambda_i z_i$ . Indeed, the mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is the constant mapping  $x \mapsto z$  if and only if  $J_A$  is the shift mapping  $x \mapsto x - z$ . Thus, we see that  $J_{\mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})} = \sum_{i \in I} \lambda_i J_{\mu A_i}$  is the shift mapping  $x \mapsto x - \mu \sum_{i \in I} \lambda_i z_i$  and, consequently, that  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is the constant mapping  $x \mapsto \sum_{i \in I} \lambda_i z_i$ . However, if we let  $A_1$  be any constant mapping,  $A_2$  be any maximally monotone mapping which is not constant,  $0 < \lambda < 1$ ,  $\lambda_1 = \lambda$  and  $\lambda_2 = 1 - \lambda$ , then  $J_{\mu A_1}$  is a shift,  $J_{\mu A_2}$  is not a shift and  $J_{\mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})} = \lambda_1 J_{\mu A_1} + (1 - \lambda_1) J_{\mu A_2}$  is not a shift. Consequently,  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is not a constant mapping. Summing up, we see that being a constant mapping is recessive w.r.t. the resolvent average.

### 4.1 Linearity and affinity

We begin our discussion with the following example:

**Example 4.2** We set  $f = \|\cdot\|$ ,  $A_1 = \partial f$  and  $A_2 = \mathbf{0}$ . Then

$$J_{A_1} x = \begin{cases} \left(1 - \frac{1}{\|x\|}\right) x, & \text{if } \|x\| > 1; \\ 0, & \text{if } \|x\| \leq 1 \end{cases}$$

and  $J_{A_2} = \text{Id}$  (see [5, Example 23.3 and Example 14.5]). We see that  $J_{A_1}$  is not an affine relation and  $J_{A_2}$  is linear. However, letting  $0 < \lambda < 1$ ,  $\lambda_1 = \lambda$  and  $\lambda_2 = 1 - \lambda$ , then

$$J_{\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})}x = \lambda J_{A_1}x + (1 - \lambda)J_{A_2}x = \begin{cases} \left(1 - \lambda \frac{1}{\|x\|}\right)x, & \text{if } \|x\| > 1; \\ (1 - \lambda)x, & \text{if } \|x\| \leq 1, \end{cases}$$

which is not an affine relation. Thus, by employing Fact 1.16, we conclude that  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$  is not an affine relation.

Example 4.2 demonstrates that linearity and affinity are not dominant properties w.r.t. the resolvent average. On the other hand, since the sums and inversions in the definition of  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  preserve linearity and affinity, we arrive at the following observation:

**Corollary 4.3 (Linearity and affinity are recessive)** *Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is a maximally monotone linear (resp. affine) relation. Then  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is a maximally monotone linear (resp. affine) relation.*

## 4.2 Rectangularity and paramonotonicity

We recall Definition 1.14 of rectangular and paramonotone mappings and consider the following example:

**Example 4.4** In  $\mathbb{R}^2$ , let  $A_1 = N_{\mathbb{R} \times \{0\}}$ . Then by Fact 1.6,  $J_{A_1}$  is the projection on  $\mathbb{R} \times \{0\}$ . Since  $A_1$  is a subdifferential of a proper, lower semicontinuous and convex function, it is rectangular (see [5, Example 24.9]) and paramonotone (see [5, Example 22.3]). Let  $A_2 : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be the counterclockwise rotation by  $\pi/2$ . Then by employing standard matrix representation we write

$$J_{A_1} = P_{\mathbb{R} \times \{0\}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad A_{2+} = \frac{1}{2}(A_2 + A_2^\top) = 0 \quad \text{and} \quad J_{A_2} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Letting  $\lambda_1 = \lambda_2 = \frac{1}{2}$ , we obtain

$$\mathcal{R}(\mathbf{A}) = \left(\frac{1}{2}J_{A_1} + \frac{1}{2}J_{A_2}\right)^{-1} - \text{Id} = \begin{pmatrix} 0 & -1 \\ 1 & 2 \end{pmatrix} \quad \text{and} \quad \mathcal{R}(\mathbf{A})_+ = \frac{1}{2}(\mathcal{R}(\mathbf{A}) + \mathcal{R}(\mathbf{A})^\top) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}.$$

By employing Fact 1.15 we see that  $A_2$  as well as  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$  are neither rectangular nor paramonotone.

In view of Example 4.4 it is clear that rectangularity and paramonotonicity are not dominant properties w.r.t. the resolvent average. We now prove the recessive nature of these properties. We begin with rectangularity.

**Theorem 4.5 (rectangularity is recessive)** *Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is rectangular and maximally monotone. Then  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is rectangular.*

*Proof.* The mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is rectangular if and only if  $\mu A$  is rectangular as can be seen from (13). Thus, by employing (14) we see that  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is rectangular as soon as  $\text{dom } \mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) \times \text{ran } \mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) \subseteq \text{dom } F_{\mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})}$ . Indeed, since for each  $i \in I$ ,  $\text{dom } \mu A_i \times \text{ran } \mu A_i \subseteq \text{dom } F_{\mu A_i}$ , by employing inclusion (31) and then inclusion (38) we arrive at

$$\begin{aligned} \text{dom } \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) \times \text{ran } \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) &\subseteq \sum_{i \in I} \lambda_i (\text{dom } \mu A_i \times \text{ran } \mu A_i) \subseteq \sum_{i \in I} \lambda_i \text{dom } F_{\mu A_i} \\ &\subseteq \text{dom } F_{\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})}, \end{aligned}$$

as asserted. ■

In order to prove that paramonotonicity is recessive, we will make use of the following result:

**Proposition 4.6** *Suppose that for each  $i \in I$ ,  $T_i : \mathcal{H} \rightarrow \mathcal{H}$  is firmly nonexpansive and set  $T = \sum_{i \in I} \lambda_i T_i$ . Then:*

(i) *If for each  $i \in I$ , given points  $x$  and  $y$  in  $\mathcal{H}$ ,*

$$\|T_i x - T_i y\|^2 = \langle x - y, T_i x - T_i y \rangle \quad \Rightarrow \quad \begin{cases} T_i x = T_i(T_i x + y - T_i y) \\ T_i y = T_i(T_i y + x - T_i x), \end{cases} \quad (51)$$

*then  $T$  also has property (51).*

(ii) *If there exists  $i_0 \in I$  such that  $T_{i_0}$  has property (51) and is injective, then  $T$  has property (51) and is injective.*

*Proof.* (i) Suppose that  $x$  and  $y$  are points in  $\mathcal{H}$  such that  $\|Tx - Ty\|^2 = \langle x - y, Tx - Ty \rangle$ . Then Corollary 1.10 guarantees that  $T_i x - T_i y = Tx - Ty$  for every  $i \in I$ . Employing property (51) of the  $T_i$ 's, we obtain the set of equalities in (51). Consequently, we see that

$$Tx = \sum_{i \in I} \lambda_i T_i x = \sum_{i \in I} \lambda_i T_i(T_i x + y - T_i y) = \sum_{i \in I} \lambda_i T_i(Tx + y - Ty) = T(Tx + y - Ty)$$

and, similarly, that  $Ty = T(Ty + x - Tx)$ . (ii) In the same manner as in the proof of (i), if  $\|Tx - Ty\|^2 = \langle x - y, Tx - Ty \rangle$ , then  $T_i x - T_i y = Tx - Ty = T_{i_0} x - T_{i_0} y$  for every  $i \in I$ . Since  $T_{i_0}$  is injective and has property (51), then, in fact,  $x - y = T_{i_0} x - T_{i_0} y$ . Thus, we conclude that  $x = Tx + y - Ty$ , equivalently  $y = Ty + x - Tx$ , which confirms that  $T$  has property (51). Finally, Proposition 1.10(i) guarantees that  $T$  is also injective. ■

**Theorem 4.7 (paramonotonicity is recessive)** *Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is maximally monotone and paramonotone. Then  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is paramonotone.*

*Proof.* First we consider the case  $\mu = 1$ . We recall that the maximally monotone mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is paramonotone if and only if  $J_A$  has property (51) (see [11, Theorem 2.1(xv)]). Thus, since each  $J_{A_i}$  has property (51), then Proposition 4.6(i) guarantees that  $J_{\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})} = \sum_{i \in I} \lambda_i J_{A_i}$

also has property (51), which, in turn, implies that  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$  is paramonotone. For arbitrary  $0 < \mu$ , we note that the mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is paramonotone if and only if  $\mu A$  is paramonotone. Since for each  $i \in I$ ,  $\mu A_i$  is paramonotone, so is  $\mathcal{R}(\mu A, \boldsymbol{\lambda})$ . Consequently, formula (26) implies that  $\mu^{-1} \mathcal{R}(\mu A, \boldsymbol{\lambda}) = \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is paramonotone.  $\blacksquare$

### 4.3 $k$ -cyclic monotonicity and cyclic monotonicity

We recall that the mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is said to be  $k$ -cyclically monotone, where  $k \in \{2, 3, \dots\}$ , if any  $k$  pairs  $(x_1, u_1), \dots, (x_k, u_k) \in \text{gra } A$ , letting  $x_{k+1} = x_1$ , satisfies  $0 \leq \sum_{i=1}^k \langle u_i, x_i - x_{i+1} \rangle$ . The mapping  $A$  is said to be cyclically monotone if it is  $k$ -cyclically monotone for every  $k \in \{2, 3, \dots\}$ . Rockafellar's well known characterization from [30] asserts that proper, maximally monotone and cyclically monotone mappings are precisely the subdifferentials of proper, convex and lower semicontinuous functions.

**Example 4.8** Let  $\mathcal{H} = \mathbb{R}^2$ . We set  $n = 2$ ,  $\boldsymbol{\lambda} = (1/3, 2/3)$ , we let  $A_1$  be the identity and we let  $A_2$  be the mapping obtained by the counter-clockwise rotation by  $\pi/2$ . Then  $A_1$  is cyclically monotone while  $A_2$  is monotone but not 3-cyclically monotone (see [2, Example 4.6] for further discussion of  $k$ -cyclic monotonicity of rotations by  $\pi/k$ ). Then, by employing matrix representation, we obtain

$$\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}) = \frac{1}{13} \begin{pmatrix} 5 & -12 \\ 12 & 5 \end{pmatrix}.$$

We now let  $x_1 = x_4 = 0$ ,  $x_2 = e_1$  and  $x_3 = e_2$  where  $(e_1, e_2)$  is the standard basis of  $\mathbb{R}^2$ . Then

$$\sum_{i=1}^3 \langle \mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}) x_i, x_i - x_{i+1} \rangle = \langle \mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}) e_1, e_1 - e_2 \rangle + \langle \mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}) e_2, e_2 \rangle = -2/13 < 0.$$

Thus, we see that  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$  is not 3-cyclically monotone.

In view of Example 4.8, the property of  $k$ -cyclic monotonicity is not dominant w.r.t. the resolvent average. In order to conclude the recessive nature of  $k$ -cyclic monotonicity and of cyclic monotonicity we recall the following: By employing cyclic monotonicity, the convexity of the set of proximal mappings was recovered in [2, Theorem 6.7]. In other words, if we suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is maximally monotone and cyclically monotone, then  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$  is maximally monotone and cyclically monotone. Since the proof of this result was actually carried out for every fixed  $k$ , it, in fact, holds also for  $k$ -cyclically monotone mappings. Summing up, we arrive at the following conclusion:

**Theorem 4.9 ( $k$ -cyclic monotonicity and cyclic monotonicity are recessive)** *Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is maximally monotone. If for every  $i \in I$ ,  $A_i$  is  $k$ -cyclically monotone, then  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is  $k$ -cyclically monotone. If for every  $i \in I$ ,  $A_i$  is cyclically monotone, then  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is cyclically monotone.*

*Proof.* The case where  $\mu = 1$  follows from our discussion above. For arbitrary  $\mu > 0$ , we note that the mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is  $k$ -cyclically monotone (cyclically monotone) if and only if the mapping  $\mu A$  is  $k$ -cyclically monotone (cyclically monotone). Thus, we see that  $\mathcal{R}(\mu \mathbf{A}, \boldsymbol{\lambda})$  is  $k$ -cyclically monotone (cyclically monotone), and, consequently, by employing formula (26), so is  $\mu^{-1} \mathcal{R}(\mu \mathbf{A}, \boldsymbol{\lambda}) = \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$ .  $\blacksquare$

#### 4.4 Weak sequential closedness of the graph

The mapping  $T : \mathcal{H} \rightarrow \mathcal{H}$  is said to be *weakly sequentially continuous* if it maps a weakly convergent sequence to a weakly convergent sequence, that is,  $x_n \rightharpoonup x \Rightarrow Tx_n \rightharpoonup Tx$ . We recall that the maximally monotone mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  has a weakly sequentially closed graph if and only if  $J_A$  is weakly sequentially continuous (see [11, Theorem 2.1(xxi)]).

**Example 4.10** Within the settings of Example 4.2, we consider the case where  $\mathcal{H} = \ell^2(\mathbb{N})$  and  $\{e_1, e_2, \dots\}$  is its standard basis. Then, for every  $n \in \mathbb{N}$ , we set  $x_n = e_1 + e_n$ . Consequently,  $x_n \rightharpoonup e_1$  and  $J_{A_1} x_n = (1 - 1/\sqrt{2})x_n \rightharpoonup (1 - 1/\sqrt{2})e_1$  while  $J_{A_1} e_1 = 0$ . Thus,  $J_{A_1}$  is not weakly sequentially continuous and  $J_{A_2}$  is. Furthermore,

$$\begin{aligned} J_{\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})} x_n &= \lambda_1 J_{A_1} x_n + \lambda_2 J_{A_2} x_n = \lambda_1 (1 - 1/\sqrt{2})x_n + \lambda_2 x_n \\ &\rightharpoonup \lambda_1 (1 - 1/\sqrt{2})e_1 + \lambda_2 e_1 = (1 - \lambda_1/\sqrt{2})e_1. \end{aligned}$$

Since  $J_{\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})} e_1 = \lambda_2 e_1$ , we see that  $J_{\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})}$  is not weakly sequentially continuous, which, in turn, implies that the graph of  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$  is not weakly sequentially closed, although, the graph of  $A_2$  is.

We see that the weak sequential closedness of the graph is not a dominant property w.r.t. the resolvent average.

**Corollary 4.11 (weak sequential closedness of the graph is recessive)** *Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is maximally monotone with a weakly sequentially closed graph. Then  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  has a weakly sequentially closed graph.*

*Proof.* Since for each  $i \in I$ ,  $J_{\mu A_i}$  is weakly sequentially continuous, so is  $J_\mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) = \sum_{i \in I} \lambda_i J_{\mu A_i}$ . Consequently,  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$  has a weakly sequentially closed graph.  $\blacksquare$

#### 4.5 Displacement mappings

The mapping  $D : \mathcal{H} \rightarrow \mathcal{H}$  is said to be a *displacement mapping* if there exists a nonexpansive mapping  $N : \mathcal{H} \rightarrow \mathcal{H}$  such that  $D = \text{Id} - N$ , in which case  $D$  is maximally monotone.

**Example 4.12** Let  $n = 2$ ,  $\lambda_1 = \lambda_2 = \frac{1}{2}$ ,  $0 < \alpha_1 = \alpha$ ,  $0 < \alpha_2 = \beta$ ,  $A_1 = \alpha \text{Id}$  and  $A_2 = \beta \text{Id}$ . Then by applying formula (47) we obtain

$$\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}) = r(\boldsymbol{\alpha}, \boldsymbol{\lambda}) \text{Id} = \frac{\alpha + \beta + 2\alpha\beta}{2 + \alpha + \beta} \text{Id}.$$

Now we let  $\alpha = 2$  and  $\beta = 5$ . Then  $A_1$  is a displacement mapping,  $A_2$  is not a displacement mapping and  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}) = 3\text{Id}$  is not a displacement mapping.

In view of Example 4.12, we see that being a displacement mapping is not a dominant property w.r.t. the resolvent average. In order to prove that being a displacement mapping is recessive we will make use of the following result:

**Proposition 4.13** *The maximally monotone mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is  $\frac{1}{2}$ -strongly monotone if and only if  $A^{-1}$  is a displacement mapping.*

*Proof.* The assertion that the mapping  $A$  is  $\frac{1}{2}$ -strongly monotone is equivalent to  $A = \frac{1}{2}\text{Id} + M_1$  for a maximally monotone mapping  $M_1$ , which, in turn, is equivalent to  $A = \frac{1}{2}\text{Id} + M_2 \circ (\frac{1}{2}\text{Id})$  for a maximally monotone mapping  $M_2$  (let  $M_2 = M_1 \circ (2\text{Id})$ ). By employing equation (18), this is equivalent to  $A^{-1} = 2[\text{Id} - (M_2^{-1} + \text{Id})^{-1}]$ . Finally, this is equivalent to  $A^{-1} = 2(\text{Id} - F)$  for a firmly nonexpansive mapping  $F$  (let  $F = J_{M_2^{-1}}$ ), which, in turn, is equivalent to  $A^{-1} = 2(\text{Id} - (\text{Id} + N)/2) = \text{Id} - N$  for a nonexpansive mapping  $N$  (let  $N = 2F - \text{Id}$ ). ■

**Corollary 4.14** *The maximally monotone mapping  $N : \mathcal{H} \rightarrow \mathcal{H}$  is nonexpansive if and only if  $N = 2J_B - \text{Id}$  for a maximally monotone and nonexpansive mapping  $B$ .*

*Proof.* Since  $\frac{1}{2}(N + \text{Id})$  is  $\frac{1}{2}$ -strongly monotone, the assertion  $(B + \text{Id})^{-1} = J_B = \frac{1}{2}(N + \text{Id})$  is equivalent to  $B + \text{Id}$  being a displacement mapping, that is  $\text{Id} + B = \text{Id} - N'$  for a nonexpansive mapping  $N'$ , that is,  $B$  is a nonexpansive mapping. ■

**Corollary 4.15 (being a displacement mapping is recessive)** *Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is a displacement mapping. Then  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$  is a displacement mapping.*

*Proof.* By employing Proposition 4.13 we see that for each  $i \in I$ ,  $A_i^{-1}$  is  $(1/2)$ -strongly monotone. Consequently, Theorem 3.20 guarantees that  $(\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}))^{-1} = \mathcal{R}(\mathbf{A}^{-1}, \boldsymbol{\lambda})$  is  $(1/2)$ -strongly monotone, which, in turn, implies that  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$  is a displacement mapping. ■

## 4.6 Nonexpansive monotone operators

In Example 4.12, we let  $\alpha = 1$  and  $\beta = 5$ . Then  $A_1 = \text{Id}$  is nonexpansive,  $A_2 = 5\text{Id}$  is not nonexpansive and  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}) = 2\text{Id}$  which is not nonexpansive. Thus, we see that nonexpansiveness is not a dominant property.

**Theorem 4.16 (nonexpansiveness is recessive)** *Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightarrow \mathcal{H}$  is a nonexpansive and monotone mapping. Then  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$  is nonexpansive. Furthermore, for each  $i \in I$ ,  $A_i = 2J_{B_i} - \text{Id}$  where  $B_i$  is maximally monotone, nonexpansive and  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}) = 2J_B - \text{Id}$  where  $B$  is the maximally monotone and nonexpansive mapping given by  $B = \sum_{i \in I} \lambda_i B_i$ .*



*Proof.* Since the  $A_i$ 's have full domain and are continuous, they are maximally monotone. Employing Corollary 4.14, we see that the  $B_i$ 's are maximally monotone and nonexpansive (and have full domain). Furthermore, we have

$$\begin{aligned} J_{\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})} &= \sum_{i \in I} \lambda_i J_{A_i} = \sum_{i \in I} \lambda_i (2J_{B_i})^{-1} = \sum_{i \in I} \lambda_i (J_{B_i})^{-1} \circ (\tfrac{1}{2} \text{Id}) = \sum_{i \in I} \lambda_i (B_i + \text{Id}) \circ (\tfrac{1}{2} \text{Id}) \\ &= (B + \text{Id}) \circ (\tfrac{1}{2} \text{Id}) = (2J_B)^{-1}. \end{aligned}$$

Thus, we see that  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}) = 2J_B - \text{Id}$ , as asserted.  $\blacksquare$

## 5 Miscellaneous observations and remarks

In our last section, we consider combinations of properties, indeterminate properties and other observations and remarks.

### 5.1 Paramonotonicity combined with single-valuedness

**Theorem 5.1 (paramonotonicity combined with single-valuedness is dominant)** *Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$  is maximally monotone. If for some  $i_0 \in I$ ,  $A_{i_0}$  is paramonotone and at most single-valued, then so is  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$ .*

*Proof.* We recall that the maximally monotone mapping  $A : \mathcal{H} \rightrightarrows \mathcal{H}$  is at most single-valued if and only if  $J_A$  is injective (see [11, Theorem 2.1(iv)]). We also recall that  $A$  is paramonotone if and only if  $J_A$  has property (51). Thus, we see that  $J_{\mu A_{i_0}}$  is injective and has property (51). Consequently, Proposition 4.6(ii) implies that  $J_{\mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})} = \sum_{i \in I} \lambda_i J_{\mu A_i}$  is injective and has property (51), which, in turn, implies that  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is paramonotone and at most single-valued.  $\blacksquare$

### 5.2 Linear relations, bounded linear operators and linear operators on $\mathbb{R}^n$

Within the class of maximally monotone and linear relation on  $\mathcal{H}$ , which we will denote by  $MLR(\mathcal{H})$ , lies the class of classical (single-valued) monotone and bounded linear operators which we will denote by  $BML(\mathcal{H})$ . Within  $BML(\mathcal{H})$  we will denote by  $BMLI(\mathcal{H})$  the class of invertible operators, that is, bounded linear monotone surjective operators with a bounded inverse. We begin our discussion of these classes of operators with the following result:

**Theorem 5.2 (in  $MLR(\mathcal{H})$ , being  $BML(\mathcal{H})$  and being  $BMLI(\mathcal{H})$  are dominant)** *Suppose that for each  $i \in I$ ,  $A_i \in MLR(\mathcal{H})$  and there exists  $i_0 \in I$  such that  $A_{i_0} \in BML(\mathcal{H})$ . Then  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) \in BML(\mathcal{H})$ . Furthermore:*

- (i) *If  $A_{i_0} \in BMLI(\mathcal{H})$ , then  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) \in BMLI(\mathcal{H})$ ;*



(ii) If  $A_{i_0}$  is paramonotone, then  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is paramonotone.

*Proof.* Corollary 4.3 guarantees that  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is a monotone linear relation. Since  $A_{i_0}$  has a full domain, then Theorem 3.1(ii) guarantees that  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  has full domain. Since  $A_{i_0}$  is single-valued, then Theorem 3.4 guarantees that  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is single-valued. Finally, since a monotone operator is locally bounded on the interior of its domain (see [31, Theorem 1]) and since  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  has full domain, then it is everywhere locally bounded, which, in the case of single-valued linear operators, is equivalent to boundedness. If  $A_{i_0}$  is a bounded linear operator which is also invertible, then  $A_{i_0}^{-1}$  is bounded, which implies that  $\mathcal{R}_{\mu^{-1}}(\mathbf{A}^{-1}, \boldsymbol{\lambda}) = (\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}))^{-1}$  is bounded. Finally, if  $A_{i_0}$  is paramonotone, then Theorem 5.1 guarantees that  $\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})$  is paramonotone. ■

**Remark 5.3** Within the settings and hypothesis of Theorem 5.2, we note that when  $\mathcal{H} = \mathbb{R}^n$ , paramonotonicity is interchangeable with rectangularity since, in this case, the two properties are equivalent (see Fact 1.15).

We now discuss the class of maximally monotone linear relations on  $\mathbb{R}^n$ , which we denote by  $MLR(n)$ , its subclass of monotone linear mappings (classical, single-valued), which we denote by  $ML(n)$ , and its subclass of monotone, linear and (classically) invertible operators which we denote by  $MLI(n)$ . We will identify operators in  $ML(n)$  with their standard matrix representation. A subclass of  $ML(n)$  is the class of positive semidefinite matrices and a subclass of  $MLI(n)$  is the class of positive definite matrices. The resolvent average of positive semidefinite and definite matrices was studied in [12]. We now recall that any positive semidefinite matrix  $A$  has the property that  $0 \leq \det A$ . This fact holds in the larger class  $ML(n)$  as well:

**Proposition 5.4** Suppose that  $A \in ML(n)$ . Then for every  $0 \leq \lambda < 1$ ,  $0 < \det(\lambda A + (1 - \lambda) \text{Id})$ . Consequently,  $0 \leq \det A$ .

*Proof.* Let  $x \in \mathbb{R}^n$ ,  $x \neq 0$ . Since  $0 \leq \langle x, Ax \rangle$ , then

$$0 < \lambda \langle x, Ax \rangle + (1 - \lambda) \|x\|^2 = \langle x, (\lambda A + (1 - \lambda) \text{Id})x \rangle.$$

We conclude that  $0 \neq (\lambda A + (1 - \lambda) \text{Id})x$  for every  $x \in \mathbb{R}^n$ ,  $x \neq 0$ . We now define  $\varphi : [0, 1] \rightarrow \mathbb{R}$  by  $\varphi(\lambda) = \det(\lambda A + (1 - \lambda) \text{Id})$ . It follows that  $\varphi$  is continuous,  $\varphi$  does not vanish on  $[0, 1[$  and  $\varphi(0) = \det \text{Id} = 1$ . Consequently,  $0 < \varphi(\lambda)$  for every  $0 \leq \lambda < 1$ , which, in turn, implies that  $\det A = \varphi(1) = \lim_{\lambda \rightarrow 1^-} \varphi(\lambda) \geq 0$ . ■

We now focus our attention on  $ML(n)$ . To this end, we first recall the case where a mapping  $A \in MLR(n)$  is, in fact, in  $ML(n)$ . Such a characterization is the combination of [14, Fact 2.2] and [14, Fact 2.3]:

**Fact 5.5** For  $A \in MLR(n)$  the following assertions are equivalent:

- (i)  $A \in ML(n)$ ;

- (ii)  $A$  is at most single-valued;
- (iii)  $A(0)$  is a singleton;
- (iv)  $\text{dom } A = \mathbb{R}^n$ .

In our current discussion it is crucial to distinguish between the notion of the inverse of a mapping in  $MLR(n)$  (multivalued settings) and the classical notion of the inverse of a single-valued mapping in  $MLI(n)$ . In the case where  $A \in MLI(n)$ , we will abuse the notation and write  $A^{-1}$  for its inverse, which can then be viewed as the same in both settings, the multivalued and the classical single-valued. In order for our results in the present paper to be relevant for studies within classical linear algebra settings in  $ML(n)$ , we now support our claim that:

When taking the resolvent average of operators in  $ML(n)$ , all of the inversion operations involved in the averaging operation are classical inversions. Furthermore, for operators in  $MLI(n)$ , all of the inverses in the formula  $(\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}))^{-1} = \mathcal{R}_{\mu^{-1}}(\mathbf{A}^{-1}, \boldsymbol{\lambda})$  are classical inverses.

Indeed, the “Furthermore” part follows from the invertibility of the  $A_i$ ’s when we recall Theorem 5.2. Now, if  $0 < \mu$  and  $A_i \in ML(n)$ , then  $\mu A_i + \text{Id}$  is invertible (since  $0 < \|x\|^2 \leq \langle x, (\mu A_i + \text{Id})x \rangle$  for every  $0 \neq x \in \mathbb{R}^n$ ). Thus, since  $A_i$  is monotone, then  $J_{\mu A_i}$  is now seen to be invertible and firmly nonexpansive, that is, for any  $x \in \mathbb{R}^n$ ,  $x \neq 0$ , we have  $0 < \|J_{\mu A_i} x\|^2 \leq \langle x, J_{\mu A_i} x \rangle$ . As a consequence,

$$\langle x, J_{\mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})} x \rangle = \sum_{i \in I} \lambda_i \langle x, J_{\mu A_i} x \rangle > 0.$$

Since this is true for every  $x \in \mathbb{R}^n$ ,  $x \neq 0$ , we conclude that  $J_{\mu \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})}$  is, indeed, invertible, as asserted.

We now focus our attention further on  $ML(n) \cap O(n, \mathbb{R})$ . Here  $O(n, \mathbb{R})$  is the group of orthogonal matrices in  $\mathbb{R}^{n \times n}$ ; in particular,  $A \in O(n, \mathbb{R}) \Rightarrow \det A = \pm 1$ . Since, as we saw in Proposition 5.4, the determinant is greater than zero in  $ML(n)$ , then, in fact,  $ML(n) \cap O(n, \mathbb{R}) = ML(n) \cap SO(n, \mathbb{R})$ , where  $SO(n)$  is the group of the special orthogonal matrices, that is, the matrices  $A \in O(n)$  such that  $\det A = 1$ . Thus,  $ML(n) \cap O(n, \mathbb{R}) = ML(n) \cap SO(n, \mathbb{R})$  can be viewed as the subset of rotations which consists of the rotations by an acute or right angles. The next result demonstrates that the resolvent average of such rotations is again such a rotation. This, of course, fails when taking the arithmetic average.

**Theorem 5.6 (being a rotation by an acute or right angle is recessive)** *Suppose that for each  $i \in I$ ,  $A_i \in ML(n) \cap O(n, \mathbb{R}) = ML(n) \cap SO(n, \mathbb{R})$ . Then  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}) \in ML(n) \cap O(n, \mathbb{R}) = ML(n) \cap SO(n, \mathbb{R})$ .*

*Proof.* By employing the inversion formula (2.2) (which, in this case, was seen to be a classical inversion), we obtain

$$(\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}))^{-1} = \mathcal{R}(\mathbf{A}^{-1}, \boldsymbol{\lambda}) = \mathcal{R}(\mathbf{A}^\top, \boldsymbol{\lambda}) = \mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})^\top.$$

That is,  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$  is orthogonal. The fact that  $\det \mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}) = 1$  follows from the monotonicity of  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$  (see Proposition 5.4).  $\blacksquare$

**Example 5.7 (generating Pythagorean triples)** When  $n = 2$ , we consider the case where  $A_1 = \text{Id}$ ,  $A_2$  is the counter clockwise rotation by  $\frac{\pi}{2}$ ,  $0 < \lambda < 1$ ,  $\lambda_1 = \lambda$  and  $\lambda_2 = 1 - \lambda$ . Then

$$\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}) = \frac{1}{\lambda^2 - 2\lambda + 2} \begin{pmatrix} \lambda(2 - \lambda) & -2(1 - \lambda) \\ 2(1 - \lambda) & \lambda(2 - \lambda) \end{pmatrix}$$

is a counter-clockwise rotation matrix by an angle of a right triangle with sides

$$a(\lambda) = \lambda(2 - \lambda), \quad b(\lambda) = 2(1 - \lambda) \quad \text{and} \quad c(\lambda) = \sqrt{a^2 + b^2} = \lambda^2 - 2\lambda + 2.$$

Now  $\lambda \mapsto \mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$  is a smooth and one-to-one curve, in particular,  $\lambda \mapsto a(\lambda)/c(\lambda)$  is a bijection of  $[0, 1]$  to itself. (In fact, for any two monotone mappings  $A_1 : \mathcal{H} \rightrightarrows \mathcal{H}$  and  $A_2 : \mathcal{H} \rightrightarrows \mathcal{H}$  such that  $A_1 \neq A_2$ ,  $\lambda \mapsto \mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$  is one-to-one). Consequently, up to rescaling, all possible right triangles can be recovered in this manner. In particular, considering all of the possible rational values of  $\lambda$ , our procedure recovers all possible *primitive Pythagorean triples*  $(a, b, c)$  (other values of  $\lambda$  should be considered as well in order to recover all possible Pythagorean triples). Indeed, letting  $\lambda = \frac{p}{q}$  where  $0 < p < q$  are natural numbers, we obtain the matrix

$$\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}) = \frac{1}{\lambda^2 - 2\lambda + 2} \begin{pmatrix} \lambda(2 - \lambda) & -2(1 - \lambda) \\ 2(1 - \lambda) & \lambda(2 - \lambda) \end{pmatrix} = \frac{1}{p^2 - 2pq + 2q^2} \begin{pmatrix} p(2q - p) & -2q(q - p) \\ 2q(q - p) & p(2q - p) \end{pmatrix}$$

which is a counter-clockwise rotation matrix by an angle of a right triangle with sides

$$a = p(2q - p), \quad b = 2q(q - p) \quad \text{and} \quad c = p^2 - 2pq + 2q^2.$$

Letting  $k = q - p$  and  $l = q$ , we obtain the formula

$$a = l^2 - k^2, \quad b = 2kl \quad \text{and} \quad c = k^2 + l^2, \tag{52}$$

which is a well known formula for generating all of the primitive Pythagorean triples and is attributed to Euclid. For further, more accurate, relations between formula (52) and primitive Pythagorean triples as well as for an extensive historical overview see [24, Section 4.2].

### 5.3 Nonexpansive monotone operators and Banach contractions

We continue our discussion of nonexpansive monotone operators. Within this class of mappings, being a Banach contraction is a dominant property:

**Theorem 5.8 (within the class of nonexpansive mappings, being a Banach contraction is dominant)** *Suppose that for each  $i \in I$ ,  $A_i : \mathcal{H} \rightarrow \mathcal{H}$  is nonexpansive and monotone. If there exists  $i_0 \in I$  such that  $A_{i_0}$  is a Banach contraction, then  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$  is a Banach contraction.*

*Proof.* [11, Corollary 4.7] asserts that given a maximally monotone operator  $B : \mathcal{H} \rightrightarrows \mathcal{H}$ , letting  $A = 2J_B - \text{Id}$ , then  $A$  is a Banach contraction if and only if  $B$  and  $B^{-1}$  are strongly monotone. We now employ the settings and the outcome of Theorem 4.16: for every  $i \in I$ ,  $A_i : \mathcal{H} \rightarrow \mathcal{H}$  is monotone and nonexpansive,  $A_i = 2J_{B_i} - \text{Id}$  for a maximally monotone and nonexpansive mapping  $B_i$  and  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}) = 2J_B - \text{Id}$  where  $B : \mathcal{H} \rightarrow \mathcal{H}$  is the monotone and nonexpansive mapping  $B = \sum_{i \in I} \lambda_i B_i$ . We also see that  $B_{i_0}$  is strongly monotone, say with  $\epsilon_{i_0}$  being its constant of strong monotonicity. Now, let  $x$  and  $y$  be points in  $\mathcal{H}$ . Then

$$\begin{aligned} \langle Bx - By, x - y \rangle &= \sum_{i \in I} \lambda_i \langle B_i x - B_i y, x - y \rangle \geq \lambda_{i_0} \langle B_{i_0} x - B_{i_0} y, x - y \rangle \\ &\geq \lambda_{i_0} \epsilon_{i_0} \|x - y\|^2 \geq \lambda_{i_0} \epsilon_{i_0} \|Bx - By\|^2. \end{aligned}$$

Thus, we see that both,  $B$  and  $B^{-1}$  are  $\lambda_{i_0} \epsilon_{i_0}$ -strongly monotone, which, in turn, implies that  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}) = 2J_B - \text{Id}$  is a Banach contraction.  $\blacksquare$

## 5.4 Projections and normal cones

We will say that a property  $(p)$  is *indeterminate* (with respect to the resolvent average) if  $(p)$  is neither dominant nor recessive.

**Example 5.9 (being a projection is indeterminate)** Let  $A_1$  and  $A_2$  be the projections in  $\mathbb{R}^2$  onto  $\mathbb{R} \times \{0\}$  and  $\{0\} \times \mathbb{R}$ , respectively. That is,

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{ and } A_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Let  $0 < \lambda < 1$ ,  $\lambda_1 = \lambda$  and  $\lambda_2 = 1 - \lambda$ . Then

$$\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}) = \begin{pmatrix} \frac{\lambda}{2-\lambda} & 0 \\ 0 & \frac{\lambda-1}{\lambda+1} \end{pmatrix},$$

which is not a projection since  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})^2 \neq \mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$ .

**Example 5.10 (being a normal cone operator is indeterminate)** Let  $f_1 : \mathbb{R}^2 \rightarrow ]-\infty, +\infty]$  be the function  $f_1 = \iota_{\mathbb{R} \times \{0\}}$  and let  $f_2 : \mathbb{R}^2 \rightarrow ]-\infty, +\infty]$  be the function  $f_2 = \iota_{\{0\} \times \mathbb{R}}$ . Let  $A_1 : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$  be the normal cone operator  $A_1 = N_{\mathbb{R} \times \{0\}} = \partial f_1$  and let  $A_2 : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$  be the normal cone operator  $A_2 = N_{\{0\} \times \mathbb{R}} = \partial f_2$ . Let  $0 < \lambda < 1$ ,  $\lambda_1 = \lambda$  and  $\lambda_2 = 1 - \lambda$ , then

$$J_{A_1} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, J_{A_2} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } \mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}) = \begin{pmatrix} \frac{1-\lambda}{\lambda} & 0 \\ 0 & \frac{\lambda}{1-\lambda} \end{pmatrix}.$$

Thus, we see that  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda}) = \partial p(\mathbf{f}, \boldsymbol{\lambda}) = \nabla p(\mathbf{f}, \boldsymbol{\lambda})$  where  $p(\mathbf{f}, \boldsymbol{\lambda}) : \mathbb{R}^2 \rightarrow \mathbb{R}$  is the parabola  $p(\mathbf{f}, \boldsymbol{\lambda})(x, y) = \frac{1-\lambda}{2\lambda} x^2 + \frac{\lambda}{2(1-\lambda)} y^2$ . Since the antiderivative of  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$  is unique up to an additive constant and since  $p(\mathbf{f}, \boldsymbol{\lambda}) \neq \iota_C$  for any subset  $C$  of  $\mathbb{R}^2$ , we see that  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$  is not of the form  $\partial \iota_C$ , that is,  $\mathcal{R}(\mathbf{A}, \boldsymbol{\lambda})$  is not a normal cone operator.

## 5.5 Lipschitz monotone operators

The purpose of this last subsection is to point out that further study is still required in order to determine if Lipschitzness is a property which is dominant, recessive or indeterminate. Within certain classes of monotone operators we do, however, have several conclusions: (i) Within the class of monotone linear relations, Theorem 5.2(i) guarantees that Lipschitzness is dominant w.r.t. the resolvent average. However, no quantitative result is currently available with regard to the Lipschitz constant. (ii) Theorem 4.16 guarantees that 1-Lipschitzness is recessive in the case where  $\mu = 1$ . However, no such result is available for other Lipschitz constants. Furthermore, within the class of nonexpansive monotone mappings, we saw that being a Banach contraction is dominant, as asserted by Theorem 5.8, still, with no quantitative information regarding the Lipschitz constant. (iii) Within the class of subdifferential operators we do have dominance of Lipschitzness w.r.t. the resolvent average with an explicit Lipschitz constant, namely, Theorem 3.25. However, as implied by Example 3.28, our explicit Lipschitz constant is not.

Let  $\alpha > 0$ . In the following example, within the class of monotone linear operators and outside the class of subdifferential operators, we take the resolvent average of two mappings with a common sharp Lipschitz constant  $\alpha$  such that their resolvent average has a sharp Lipschitz constant  $\alpha^2$ . This is in stark contrast to the constant in Theorem 3.25 for subdifferential operators.

**Example 5.11** Let  $A_1 = \begin{pmatrix} 0 & -\alpha \\ \alpha & 0 \end{pmatrix} = A_2^\top$ . Then

$$J_{A_1} = \frac{1}{\alpha^2 + 1} \begin{pmatrix} 1 & \alpha \\ -\alpha & 1 \end{pmatrix} = J_{A_2}^\top \quad \text{and} \quad \mathcal{R}(\mathbf{A}) = \begin{pmatrix} \alpha^2 & 0 \\ 0 & \alpha^2 \end{pmatrix}.$$

We see that, indeed, the sharp Lipschitz constant of  $A_1$  and  $A_2$  is  $\alpha$  and the sharp Lipschitz constant of  $\mathcal{R}(\mathbf{A})$  is  $\alpha^2$ .

## Acknowledgments

Sedi Bartz was supported by a postdoctoral fellowship of the Pacific Institute for the Mathematical Sciences and by NSERC grants of Heinz Bauschke and Xianfu Wang. Heinz Bauschke was partially supported by the Canada Research Chair program and by the Natural Sciences and Engineering Research Council of Canada. Sarah Moffat was partially supported by the Natural Sciences and Engineering Research Council of Canada. Xianfu Wang was partially supported by the Natural Sciences and Engineering Research Council of Canada.

## References

- [1] A. Auslender and M. Teboulle, *Asymptotic Cones and Functions in Optimization and Variational Inequalities*, Springer-Verlag, 2003.

- [2] S. Bartz, H.H. Bauschke, J.M. Borwein, S. Reich and X. Wang, “Fitzpatrick functions, cyclic monotonicity and Rockafellar’s antiderivative”, *Nonlinear Analysis: Theory, Methods and Applications* 66, pp. 1198–1223, 2007.
- [3] H.H. Bauschke and J.M. Borwein, “Legendre functions and the method of random Bregman projections”, *Journal of Convex Analysis* 4(1), pp. 27–67, 1997.
- [4] H.H. Bauschke, J.M. Borwein and X. Wang, “Fitzpatrick functions and continuous linear monotone operators”, *SIAM Journal on Optimization* 18, pp. 789–809, 2007.
- [5] H.H. Bauschke and P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2011.
- [6] H.H. Bauschke, R. Goebel, Y. Lucet and X. Wang, “The proximal average: basic theory”, *SIAM Journal on Optimization* 19, pp. 766–785, 2008.
- [7] H.H. Bauschke, W.L. Hare and W.M. Moursi, “On the range of the Douglas-Rachford operator”, *preprint*, 2014. <http://arxiv.org/abs/1405.4006>
- [8] H.H. Bauschke, Y. Lucet and M. Trienis, “How to transform one convex function continuously into another”, *SIAM Review* 50, pp. 115–132, 2008.
- [9] H.H. Bauschke, E. Matoušková and S. Reich, “Projection and proximal point methods: convergence results and counterexamples”, *Nonlinear Analysis: Theory, Methods and Applications* 56(5), pp. 715–738, 2004.
- [10] H.H. Bauschke, S.M. Moffat and X. Wang, “Near equality, near convexity, sums of maximally monotone operators, and averages of firmly nonexpansive mappings”, *Mathematical Programming* 139, pp. 55–70, 2013.
- [11] H.H. Bauschke, S.M. Moffat and X. Wang, “Firmly nonexpansive mappings and maximally monotone operators: correspondence and duality”, *Set-valued and Variational Analysis* 20, pp. 131–153, 2012.
- [12] H.H. Bauschke, S.M. Moffat and X. Wang, “The resolvent average for positive semidefinite matrices”, *Linear Algebra and Its Applications* 432, pp. 1757–1771, 2010.
- [13] H.H. Bauschke and X. Wang, “The kernel average for two convex functions and its applications to the extension and representation of monotone operators”, *Transactions of the American Mathematical Society* 361, pp. 5947–5965, 2009.
- [14] H.H. Bauschke, X. Wang and L. Yao, “Rectangularity and paramonotonicity of maximally monotone operators”, *Optimization* 63, pp. 487–504, 2014.
- [15] J.M. Borwein, “Fifty years of maximal monotonicity”, *Optimization Letters* 4, pp. 473–490, 2010.
- [16] J.M. Borwein and J.D. Vanderwerff, *Convex Functions: Constructions, Characterizations and Counterexamples*, Cambridge University Press, 2010.

- [17] H. Brézis and A. Haraux, “Image d’une somme d’opérateurs monotones et applications”, *Israel Journal of Mathematics* 23, pp. 165–186, 1976.
- [18] R.S. Burachik and A.N. Iusem, *Set-Valued Mappings and Enlargements of Monotone Operators*, Springer-Verlag, 2008.
- [19] R. Cross, *Multivalued Linear Operators*, Marcel Dekker, 1998.
- [20] J. Eckstein and D.P. Bertsekas, “On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators”, *Mathematical Programming Series A* 55, pp. 293–318, 1992.
- [21] S. Fitzpatrick, “Representing monotone operators by convex functions”, *Workshop/Miniconference on Functional Analysis and Optimization (Canberra 1988)*, Proceedings of the Centre for Mathematical Analysis, Australian National University vol. 20, Canberra, Australia, pp. 59–65, 1988.
- [22] R. Goebel, W. Hare and X. Wang, “The optimal value and optimal solutions of the proximal average of convex functions”, *Nonlinear Analysis* 75, pp. 1290–1304, 2012.
- [23] N. Ghoussoub, *Self-dual Partial Differential Systems and Their Variational Principles*, Springer, 2009.
- [24] R. Laubenbacher and D. Pongelley, *Mathematical Expeditions: Chronicles by the Explorers*, Springer-Verlag, New York, 1999.
- [25] G.J. Minty, “Monotone (nonlinear) operators in Hilbert spaces”, *Duke Mathematical Journal* 29, pp. 341–346, 1962.
- [26] S.M. Moffat, *The Resolvent Average : An Expansive Analysis of Firmly Nonexpansive Mappings and Maximally Monotone Operators*, PhD Dissertation, The University of British Columbia, Okanagan, 2015. <http://circle.ubc.ca/handle/2429/51593>
- [27] J.-J. Moreau, “Proximité et dualité dans un espace hilbertien”, *Bulletin de la Société Mathématique de France*, 93, pp. 273–299, 1965.
- [28] T. Pennanen, “On the range of monotone composite mappings”, *Journal of Nonlinear and Convex Analysis* 2(2), pp. 193–202, 2001.
- [29] S. Reich, “A limit theorem for projections”, *Linear and Multilinear Algebra*, 13, pp. 281–290, 1983.
- [30] R.T. Rockafellar, “Characterization of the subdifferentials of convex functions”, *Pacific Journal of Mathematics* 17, pp. 497–510, 1966.
- [31] R.T. Rockafellar, “Local boundedness of nonlinear, monotone operators”, *Michigan Mathematical Journal* 16, pp. 397–407, 1969.
- [32] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

- [33] R.T. Rockafellar and R.J-B Wets, *Variational Analysis*, Springer-Verlag, 1998.
- [34] S. Simons, *From Hahn-Banach to Monotonicity*, Lecture Notes in Mathematics, Vol. 1693, Springer-Verlag, 2008.
- [35] Y. Yu, “Better approximation and faster algorithm using the proximal average”, *Advances in Neural Information Processing Systems 26*, (*NIPS 2013*), C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani and K.Q. Weinberger (editors), Curran Associates, Inc. pp. 458–466, 2013.
- [36] X. Wang, “Self-dual regularization of monotone operators via the resolvent average”, *SIAM Journal on Optimization* 21, pp. 438–462, 2011.
- [37] C. Zălinescu, *Convex Analysis in General Vector spaces*, World Scientific Publishing Co, 2002.